International Journal of Engineering &Technology, 7 (1.7) (2018) 137-141

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Brain tumor prediction using naïve Bayes' classifier and decision tree algorithms

Danda Shashank Reddy[1], Chinta Naga Harshitha[2], Carmel Mary Belinda[3*]

[1,2] Student,[3] Associate Professor, Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,Avadi, Chennai-62, TamilNadu, India
**Corresponding author E-mail:carmelbelinda@gmail.com*

Abstract

Now a day's many advanced techniques are proposed in diagnosing the tumor in brain like magnetic resonance imaging, computer tomography scan, angiogram, spinal tap and biospy. Based on diagnosis it is easy to predict treatment. All of the types of brain tumor are officially reclassified by the World Health Organization. Brain tumors are of 120 types, almost each tumor is having same symptoms and it is difficult to predict treatment. For this regard we are proposing more accurate and efficient algorithm in predicting the type of brain tumor is Naïve Bayes' classification and decision tree algorithm. The main focus is on solving tumor classification problem using these algorithms. Here the main goal is to show that the prediction through the decision tree algorithm is simple and easy than the Naïve Bayes' algorithm.

Keywords: D*iagnosis; Treatment; Efficient; Prediction; Proposing.*

1. Introduction

The flock(group) of abnormal cells which are augmenting in the brain is described as brain tumor. Irrespective of age, brain tumor is found to be present in any person. It may change depending upon the treatment given to a person that is from one treatment session to the other but the adverse effects caused by this will vary from one individual to the other. Brain tumors are appeared to be in different image intensities at any location and are of variety of shapes and the sizes. They can be malignant or benign. Benign brain tumors are of homogeneous structure and they are not comprised of any cancer cells. They are not only examined but also demolished surgically. Whereas, the malignant brain tumors comprises of cancer cells and are homogeneous in nature. Based on certain conditions we can predict which type of tumor is going to be occurred.

Though treatment of these various types of tumors can be easily done by several therapy like chemotherapy, radiotherapy or sometimes a combination therapy, but are always life threatening. So, diagnosing of the brain tumors is very essential for carrying the treatment forward. In recent years, the imaging tools are responsible for the significant advancement in neurology and basic Neuroscience that enabled in the vivo monitoring of the brain. MRI has been proved as one of the most powerful and versatile brain imaging modalities which allows non-invasive longitudinal and the 3-D assessment of tissue function, physiology and morphology. The information which is provided by MRI had critically increased the knowledge of normal and diseased anatomy for medical research, and plays a major role in diagnosis and treatment planning.MR imaging is the current prevailing method of choice for early recognition of brain tumor in human brain. The interpretation of MRIis basically based on the radiologist's view.

In this paper we will discuss about the data mining algorithms which are used to predict the type of tumor that occurs based upon symptoms or treatment and from the origination.Here we use mostly the decision tree algorithm and the naïve bayes classifier algorithm. The main objective is to 1.Estimate the occurence of the tumor 2.Identify the symptoms and the treatment necessary for the tumor.

Table 1: Tumor Dataset

	Tumor	**Origination**	**Symptoms**	**Treatment**	**Occurence**
0	Secondary tumor	Lungs	Cough	Selfcare	Yes
1	Secondary tumor	Kidney	Headache	Selfcare	Yes
2	Secondary tumor	Lungs	Quick growth	Surgery	No
3	Secondary tumor	Brain parts	Cough	Medications	No
4	Glioma	Glial cells	Memory loss	Medications	Yes
5	Glioma	Glial cells	Memory loss	Radiation	Yes
6	Astrocytoma	Brain parts	Loss of apetite	Surgery	Yes
7	Glioma	Stomach	Memory loss	Surgery	Yes
8	Olingodendro glioma	Brain parts	Quick growth	Radiation	Yes
9	Anaplastic astrocytoma	Brain parts	Quick growth	Surgery	No
10	Astrocytoma	Kidney	Loss of apetite	Selfcare	No
11	Secondary tumor	Kidney	Headache	Medications	No
12	Anaplastic astrocytoma	Glial cells	Quick growth	Surgery	Yes
13	Olingodendro glioma	Stomach	Memory loss	Radiation	No
14	Glioma	Glial cells	Headache	Surgery	No

Copyright © 2018S Alex David, S Ravikumar, K Antony Kumar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2. Decision Tree Algorithm

Decision tree is the learning algorithms that construct a classification tree to classify the data and it represents the visual presentation of a problem .It decompose a complex problem into several sub problems with smaller size. The process is restated until the complete tree is build. The tree is originated at the root with the selected attributes. It is based on the approach of divide and conquers.The decision can be made in structured and formal approach. The main objective is to find the variable threshold pair which best split the observations into subgroups.

The basic philosophy of all the classifiers based on decision trees is alike, there are several other possibilities for its construction.

2.1 Constituents of a decision tree:

1.Root node: It is the top most node and gives rise to several other nodes.
2.Internal nodes are the nodes with 2 or more outgoing edges and one incoming edge.
3. Leaf node: it doesn't have any outgoing edge and it is the terminal edge.

As here the decision tree has been constructed, we check the test condition starting from the root-node and allocate the switch easily to one of the outgoing edges, and at last the node is assigned after testing the condition again. After all the test conditions lead to the leaf node then the decision tree is said to be complete. The leaf node comprises the class-labels, which vote in-favor or always in contradiction of the decision.

Now, you might think why we took the 'tumor' attribute at the root? If you choose other attribute, the decision tree i.e being erected will be different. There are several other approaches for constructing a decision tree. Now we use the Greedy Approach and construct the decision tree in the following way:

2.2 The Greedy Approach

"Greedy Approach is purely grounded on the heuristic Problem Solving concept by making optimal local choice at each and every node. By doing this we get the approximate optimal solution universally."

We summarize the algorithm as :
1. first élite out the best feature as the test condition at each stage.
2. divide the nodes into the possible outcomes (internal nodes).
3. Now always recurrence the above steps for exhausting last test conditions into leaf nodes.
The first question when you start the implementation of algorithm is: 'how to pick the best starting condition?"
The best solution is it depends on the values of the entropy and the information gain.

2.2.1 Entropy

Entropy is nothing but the homogeneity. For homogenous data, the entropy is 0, else if the data is separated into half-half then it is 1.

2.2.2 Information gain

when the node is split the information gain is the increment/decrement in entropy. There should be the maximum information gain that should be nominated for piercing for the attribute. We pick the finest attribute at any particular step based on the computed values of information gain and the entropy. Here from the above dataset the decision tree can be implemented in the below following way. This can be drawn as

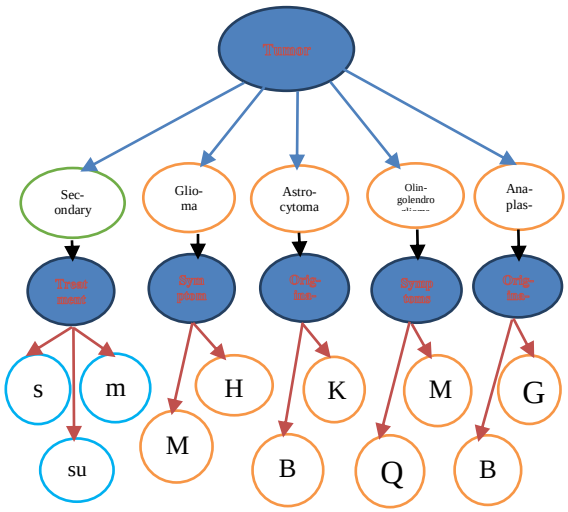


Fig1: decision tree for data set

3. Naïve Bayes'Classifier

The classifier known as Naïve Bayes Classifier is well-defined as the collection of several algorithms that are classified based on Bayes' Theorem. It is the grouping of several algorithms but not a sole algorithm where all of them always being shared with a mutual value, i.e. each couple of landscapes being confidential is always autonomous of each other. To flinch with, we will deliberate a dataset mentioned above. Consider a imagined dataset that defines the tumor symptoms,origination and treatment for the occurence of the tumor. Given the tumor dataset, every tuple categorizes the actions as if it is fit then it shows "yes" else it shows "no" for the occurence of tumor

Table 2 Tumor Dataset

	Tumor	Origi-na-tion	Symp-toms	Treatment	Oc-curence
0	Secondary tumor	Lungs	Cough	Selfcare	Yes
1	Secondary tumor	Kidney	Headache	Selfcare	Yes
2	Secondary tumor	Lungs	Quick growth	Surgery	No
3	Secondary tumor	Brain parts	Cough	Medica-tions	No
4	Glioma	Glial cells	Memory loss	Medica-tions	Yes
5	Glioma	Glial cells	Memory loss	Radiation	Yes
6	Astrocytoma	Brain parts	Loss of apetite	Surgery	Yes
7	Glioma	Stomach	Memory loss	Surgery	Yes
8	Olingodendro glioma	Brain parts	Quick growth	Radiation	Yes
9	Anaplastic astrocytoma	Brain parts	Quick growth	Surgery	No
10	Astrocytoma	Kidney	Loss of apetite	Selfcare	No
11	Secondary tumor	Kidney	Headache	Medica-tions	No
12	Anaplastic astrocytoma	Glial cells	Quick growth	Surgery	Yes
13	Olingodendro glioma	Stomach	Memory loss	Radiation	No
14	Glioma	Glial cells	Headache	Surgery	No

The above dataset will be separated into two types and they are known as **feature matrix** and the **response vector**.

- The first type i.e the feature matrix holds each and every vector i.e rows of the given dataset where each and every vector persists the value of dependent features. In above dataset, features are 'Tumor', 'Origination', 'Symptoms' and 'Treatment'.
- Response vector contains the value of class variable(prediction or output) for each row of feature matrix.
- In the overhead dataset, the name of the class variable is 'Tumor Occurence'

3.1 Assumption

The main basic Naïve Bayes hypothesis is known as each and every feature varieties an:

- independent
- equal

contribution with a required outcome.

With matching to the above dataset, we can easily classify this as:

At first we undertake that any pair is reliant on with each other. For example, the origination is from 'kidneys' has no work to do with the symptoms or if the tumor is 'astrocytoma' has no effect on the treatment. Hence, the features here are always known to be not dependent to each other. And next each and every feature is possessed with the same weight(or importance). For example, knowing only origination and symptoms alone can't predict the outcome accurately. None of the attributes is irrelevant and expected to be underwriting **similarly** to the conclusion.
Note: We can say here that the following assumptions that we made by the Naïve bayes classifier is not always correct in the real time situations. So we can say that the independent assumptions are not correct always but it works well for the purpose of practice. Now, we first discuss about the Bayes' theorem.

3.2 Bayes' Theorem

This theorem treasures the probability of a particular event that is occurring given the probability of the another event that has already occurred. This theorem can be expressed mathematically as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Here in the above equation A and B are the events and P(B)? 0.

Mostly, here we are annoying to treasure the probability of the event A, hence given that always the event B is true. EventB is also labelled as **evidence**. The prior of A has P(A) i .e (the prior probability, i.e. Probability of event before indication is seen). The indication is continuously an attribute value of an strange occurrence(here, it is event B). P(A|B) is a posteriori probability of B, i.e. probability of event after indication is seen.
Now, based on our above dataset, we use Bayes theorem in the following way:

$$P(y\,|X) = \frac{P(X\,|y)\,P(y)}{P(X)} \quad (2)$$

where, here the class variable is y and the vector i.e featured is Xi.e dependent (of size *n*) where:

$$X = (x_1, x_2, \ldots\ldots, x_n) \quad (3)$$

Just to clear, an example of a feature vector and corresponding class variable can be: (refer 1st row of dataset)

X = (Secondary Tumor, Lungs ,Cough, Self care)

y = Yes

So essentially, P(X|y) here incomes, the probability of "Occurence of tumor" given that the conditions are "tumor is secondary", "origination is lungs", "symptom is cough" and "treatment is selfcare".

3.3 Naive assumption

Now, we will put the following naïve assumption to the above Bayes theorem, which is, **independence** midst the features. So we riven the evidence into several not reliant on parts.
Now, if here any two events A and B are autonomous, then,

P(A,B) = P(A). P(B) (4)

Hence, we reach to the result:

$$P(y|x_1, \ldots\ldots, x_n) = \frac{P(x_1|y)P(x_2|y)\ldots\ldots\ldots P(x_n|y)P(y)}{P(x_1)P(x_2)\ldots\ldots\ldots P(x_n)} \quad (5)$$

which can be expressed as:

$$P(y|x_1, \ldots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)\ldots\ldots P(x_n)} \quad (6)$$

Since here the denominator ruins constant for a given particular input we can remove the term.

$$\mathbf{P(\,y\mid x_1, \ldots\ldots\ldots, x_n)} \propto \boldsymbol{P(y)\prod_{i=1}^{n} P(x_i\,|y)} \quad (7)$$

Then, we should create a classifier model. For this, we treasure the probability of given set of inputs for all possible values of the class variable *y* and élite up the output with extreme probability. This can be stated mathematically as:

$$\hat{y} = argmax_y P(y)\prod_{i=1}^{n} P(x_i\,|y) \quad (8)$$

So, in conclusion, we are left with the task of scheming P(y) and $P(x_i \mid y)$.

Please note that P(y) is also called **class probability** and $P(x_i \mid y)$ is called **conditional probability**.
The diverse naive Bayes classifiers are largely diverge chiefly by the molds they make concerning the distribution of $P(x_i \mid y)$.
Now we should try this by applying the above formula automatically on our above tumor dataset. In-order to perform this, we need to do some precomputations on our dataset.
The other tables are classified as follows:
We want to treasure $P(x_i \mid y_j)$ for each x_i in X and y_j in y.
All these intentions have been proved in the tables below:

Table 3 (type of tumor)

Tumor	Yes	No	P(Yes)	P(No)
Secondary tumor	2	3	2/8	3/7
Glioma	3	1	3/8	1/7
Astrocytoma	1	1	1/8	1/7
Olingodendro glioma	1	1	1/8	1/7
Anaplastic astrocytoma	1	1	1/8	1/7
	8	7	Total= 100%	Total= 100%

Table 4 (origination)

	Occurence	P(Yes)/P(No)
Yes	8	8/15
No	7	7/15
	15	Total = 100%

Table 5 (symptoms)

Origination	Yes	No	P(Yes)	P(No)
Lungs	1	1	1/8	1/7
Kidneys	1	2	1/8	2/7
Glial cells	3	1	3/8	1/7
Stomach	1	1	1/8	1/7
Brain parts	2	2	2/8	2/7
	8	7	Total = 100%	Total = 100%

Table 6 (treatment)

Symptoms	Yes	No	P(Yes)	P(No)
Cough	1	1	1/8	1/7
Headache	1	2	1/8	2/7
Memory loss	3	1	3/8	1/7
Loss of apetite	1	1	1/8	1/7
Quick growth	2	2	2/8	2/7
	8	7	Total = 100%	Total = 100%

Table 7 **(occurence)**

Treatment	Yes	No	P(Yes)	P(No)
Self care	2	1	2/8	1/7
Surgery	3	3	3/8	3/7
Medications	1	2	1/8	2/7
Radiation	2	1	2/8	1/7
			Total = 100%	Total = 100%

So, in the figure overhead, we have intended P(x_i | y_j) for each x_i in X and $y_j$in y manually in the tables 1-4. For example, probability of occurring of tumor given that the origination is kidney, i.e P(origin = kidney | occurence = Yes) = 1/8.
Also, we need to find class probabilities (P(y)) which has been intended in the above tables. For example, P(occurence = Yes) = 8/15.

Now here we are ready with all the assumptions here:

Let us test it on a new set of features (let us call it most occuring):

Most occuring = (glioma,lungs, memory loss, surgery)

So, probability of occurring of tumoris given by:

P(Yes|currentoccurance)=

$$\frac{P(GliomaTumour|Yes)P(LungsOrigination|Yes)P(Memorylosssymptoms|Yes)P(SurgeryTreatment|Yes)P(Yes)}{P(current\ occurance)} \quad (9)$$

and probability to not occurring of tumor is given by:

P(No|currentoccurance)=

$$\frac{P(GliomaTumour|no)P(LungsOrigination|no)P(Memorylosssymptoms|no)P(SurgeryTreatment|no)P(no)}{P(current\ occurance)} \quad (10)$$

Since, P(most occuring) is mutual in both probabilities, we can floutP(most occuring) and find proportional probabilities as:

$$P(\text{Yes}|\text{currentoccurence}) \propto \frac{3}{8}.\frac{1}{8}.\frac{3}{8}.\frac{3}{8}.\frac{8}{15} \approx 0.003515 \quad (11)$$

and

$$P(\text{No}|\text{currentoccurence}) \propto \frac{1}{7}.\frac{1}{7}.\frac{1}{7}.\frac{3}{7}.\frac{7}{15} \approx 0.000583 \quad (12)$$

Now, since

P(Yes|current occurence) + P(No|current occurence) =1

These numbers can be straight forwardly transformed to a probability by assembling the sum equal to 1 (normalization)

$$P(\text{Yes}|\text{current occurence}) = \frac{0.003515}{0.003515+0.000583} = 0.859 \quad (13)$$

and

$$P(\text{No}|\text{current occurence}) = \frac{0.000583}{0.003515+0.000583} = 0.141 \quad (14)$$

Since P(Yes|current occurence) > P(No|current occurence)

So, prediction that tumor occurs is 'Yes'.

The method that we discussed above is valid for discrete data. In case of continuous data, we need to make some expectations concerning the distribution of values of each feature. Generally there are dissimilar types of naïve Bayes classifiers and they fluctuate mainly from the assumptions they make based on the distribution of $P(x_i|y)$.

4. Conclusion

I here by conclude that many data mining techniques are used for the detection of brain tumor type among several kinds of patients. In this paper we used two classification techniques in data mining to predict the type of brain tumor in patients: Naïve Bayesian, decision tree. Using these two algorithms the type of tumor has been found and it allows analysis of historical data from data sets which helps neurologists to predict the type of tumor. From the above prediction analysis we have found that the decision tree predicts more fastly and accurately than that of the naïve Bayes classifier.

References

[1] [1] Janki naik 1, Sagar Patel 2 "Tumor Detection and Classification using Decision Tree in Brain MRI"IJCSNS (International journal of computer science and Network Security, Vol. 14 no.6, June 2014, pp.87-91.

[2] Varun Jain Sunila Gondara, (june2017),Comparative Study of Data Mining Classification Methods in Brain Tumor Disease Detection.IJCSC, Vol.8, Issue-2 pp.12-17.

[3] Kalyani A.Bhawar, Prof.Ajay S.chhajed "Brain Tumor classification using Data mining algorithms.IJESRT, November-2016, pp.239-243.

[4] V.Vani, M.Kalaiselvi Geetha Automatic Tumor Classification of Brain MRI Images .IJCSC, Vol-4, Issue-10, pp.144-151.

[5] Mr. Meena, E Murali Study on Various Machine Learning Algorithms for Brain Tumor Detection.Ijpam, Volume-117 No.8 2017, pp.139-143.
[6] Chau, M.; Shin, D. (2009). A Comparative Study of Medical Data Classification Methods Based onDecision Tree and Bagging Algorithms. Proceedings of IEEE International Conference onDependable, Autonomic and Secure Computing, pp. 183-187.
[7] Kearns M. and Mansour Y. On the boosting ability of top-down decision treelearning algorithms.Journal of Computer and Systems Sciences, 58(1): 109-128, 1999.
[8] Brain Tumor ppt. Dr.Walaa Nasr Lecturer of Medical-Surgical Nursing Department Second year 2012.