Supervised Learning Algorithms for Detection of Brain Tumour

Pranitha Bhuta, Konda HimaKireeti, Nazma Mohammad

Abstract : Over the past few years in the field of medicine, data mining is being used for the prediction of diseases. Data mining is the technique of extracting significant data from massive warehouses or repositories or other datasets. Brain tumour is inherently serious and fatal due to its behaviour in the confined space of the intracranial cavity. Several sophisticated methods are being used in detecting the brain tumour such as Biopsy, Angiogram, Magnetic Resonance Imaging and Spinal Tap. The treatment can be predicted by proper diagnosis in early stages. World Health Organization reclassified all types of brain tumour officially. There are 120 types of Brain tumours, having similar symptoms and hence the treatment cannot be predicted easily. Studies have determined that most of people with brain tumours have died as a result of inaccurate detection. To overcome this, in this paper we have proposed an effective and precise algorithm that predicts the type of brain tumour. Algorithms such as Decision Tree and Naïve Bayes' classification are chosen. The focus of this paper is how these algorithms would classifying the types of brain tumours with ease and accuracy.

Keywords :Entropy, gain , decision tree, symptoms, classifier.

I. INTRODUCTION

Brain tumour refers to a congregation of anomalous cells which grow in the brain. It can be found in individuals of any age. The characteristics of the tumour change according to treatment the person undergoes which may result in adverse effects that vary from individual to individual. The tumours at any random site appearin varying image intensities and are of distinct sizes and shapes. Tumours are either benign or malignant. Homogeneous structured tumours which are benign are free of any cancer cells. After being examined they are surgically annihilated. When it comes to malignant brain tumours, they are comprised of cancer cells and in nature are homogeneous[1]. Certain conditions let us predict the tumour type that is going to occur. A person who is already affected by any kind of tumour is susceptible to brain tumour of any type. Though treatment of those varied sorts of tumours may be simply done by many medical aids like therapy actinotherapy or radiotherapy but continue to be life threatening. Assessment of brain tumours is very important for any further diagnosis.

Revised Manuscript Received on June 05, 2019

Pranitha Bhuta, Department of Computer Science, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.

Konda Himakireeti, Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.

Nazma Mohammad, Department of Computer Science, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.

Over the years, the imaging tools are the reason for numerous advancements in neuroscience and rudimental neurobiology that enable the detailed study of the brain. [5] Later , MRI evolved as the foremost powerful, as well as most advanced brain imaging modalities that permits the assessment of morphology in 3-D, functioning of tissues and physiology. Data that is obtained by MRI had critically augmented the information of traditional and infected anatomy for medical analysis, and takes a vital responsibility in identification and diagnosis.MR imaging is the presently used methodology of preference for advance recognition of neoplasm in human brain. The MRI interpretation is primarily based on the perspective of a radiologist. The data mining algorithms discussed in this paper, are accustomed to predict the kind of tumour that happens primarily based upon treatment, symptoms and their origin. The primary objective is the estimation of the occurrence of the tumour and to identify the symptoms , along with the treatment needed.

II. DECISION TREE ALGORITHM

It is an algorithm that constructs any tree of classification to classify the information and gives the pictorial representation of an issue[8]. Decomposes a complicated issue into several sub issues of smaller sizes. This method is repeated till the whole tree is constructed. The tree begins with a root which is a particular selected attribute [2][3]. The approach of divide and conquer is followed. The primary aim is always to search the variable threshold pair that divides the subgroups in a best possible way. The basic idea of classifications derived from decision trees are similar, there are many different potentialities for its construction.

2.1 Constituents of a decision tree

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

2.2 Entropy

Entropy is defined as the sum of each label's probability times the same label's logarithmic probability. It is simply the homogeneity. Homogenous data has the entropy of 0, and for the data that is split into two equal halves, it is 1.

2.3 Information gain

On splitting, the node has some increment/decrement in entropy which is the information gain. [1]For piercing, the attribute with maximum information gain is selected. Based on the calculated values of entropy and information gain the finest attribute at every step is selected. In

this paper, for the above sample dataset which is taken from our dataset containing 270 observations(sample shown table-1), the decision tree is constructed. To construct a decision tree, we should first find out the entropies of all the attributes present in the table(table-1). We have the attributes Tumour, Symptom, Treatment, Origination. To find the entropy, information gain is needed. Formula to calculate information gain is

I(y,n)= -y/(y+n) log(y/y+n)-n/(y+n) log(y/y+n).

Where y is the no. of yes's in the table and n is the no. of no's in the table.

Therefore, y=7 and n=8.

IG(y,n)=0.995

Calculate the entropy of tumour now.

E(Attribute)= $\sum_{i=1}^{n}((y_i + n_i)/(y+n)) * IG(y_i, n_i)$

Table-2

Tumour	y_i	n_i	$IG(y_i, n_i)$
Oligodendrogliomas	1	2	0.914
Secondary	2	1	0.914
Medulloblastomas	1	2	0.914
Glioma	2	1	0.914
Astrocytoma	1	2	0.914

Entropy (TUMOR)=0.5484

GAIN(TUMOR)=E-IG=0.995-0.5484=0.447

Similarly,

GAIN(ORIGIN)=0.114

GAIN(SYMPTOMS)=0.356

GAIN(TREATMENT)=0.039

	Tumour	Symptom	Originat-ion	Treatment	Occu-rrence
1	Oligodendr-ogliomas	Seizures	Lungs	Radiation	No
2	Secondary	Headache	Lungs	Medicatio-ns	Yes
3	Medullob-lastomas	Nausea	Lungs	Surgery	No
4	Secondary	Headache	Kidney	Medicatio-ns	Yes
5	Oligoden-drogliomas	Nausea	Brain parts	Radiation	Yes
6	Glioma	Memory-Loss	Stomach	Surgery	Yes
7	Medullob-lastomas	Nausea	Kidney	Radiation	No
8	Glioma	Seizures	Glial cells	Surgery	No
9	Oligoden-drogliomas	Memory-Loss	Stomach	Surgery	No
10	Glioma	Memory-Loss	Glial cells	Radiation	Yes
11	Medullob-lastomas	Headache	Brain parts	Surgery	Yes
12	Secondary	Nausea	Brain parts	Medicatio-ns	No
13	Astrocytoma	Seizures	Brain parts	Radiation	Yes
14	Astrocytoma	Nausea	Lungs	Surgery	No
15	Astrocytoma	Seizures	Stomach	Medicatio-ns	No

Tumour Table (Table-1)

As tumour has the maximum gain value, it will become the root node for our decision tree as shown in our fig-1.

Following similar process and calculating the entropies and gain at every level for every node(attribute), the following decision tree is constructed(fig-1):

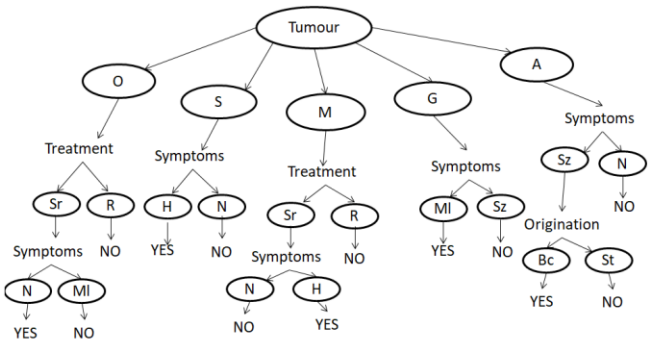

Fig-1

DECISION TREE

Abbreviations are as follows:

O = Oligodendrogliomas Tumour
S = Secondary Tumour
M = Medulloblastomas Tumour
G = Glioma Tumour
A = Astrocytoma Tumour
Sr = Surgery
R = Radiation
H =Headache
N = Nausea
Ml = Memory loss
Sz = Seizure
Bc = Brain cells
St = Stomach

III. NAÏVE BAYES'CLASSIFIER

Among the collection of algorithms which are based on Bayes' Theorem, the Naïve Bayes Classifier is the most significant classifier. This classifier is coercing of various algorithms but not just a single algorithm where all of those algorithms are always being shared with a common value, i.e. each pair of landscapes being confidential are autonomous to one other [9]. Now to experiment with, we will circumspect the dataset shown in the above table (see table-1). Let's consider this imagined dataset that describes the tumour, origination, symptoms, and treatment showing the occurrence of that tumour. In any given tumour dataset, each tuple categorizes the results, if it fits then "yes" is displayed, else "no" is displayed for the tumour occurrence. The dataset in the above table(table-1) will be divided into two types, called feature matrix and response vector[6].

• The feature matrix, being the primary type holds every vector meaning every row of the given dataset in whicheach vector holds the value of a dependent feature. In our dataset, the features are 'Tumour', 'Symptom' 'Origination' and 'Treatment'.

• The second type, the response vector consists of the value of a class variable(output/prediction) for every tuple inthe feature matrix.

• In our dataset, the name of the class variable will be 'Occurrence'.

3.1 Assumption

The primary hypothesis by Naïve Bayes states that all the features varieties an individualisticas well as uniform contribution to the necessary result. [4]

The classification, with respect to our dataset, is as follows: In the beginning, we take into consideration that all pairs are in reliance with

each other. For instance, the origin being from 'lungs' do not have regard with the symptoms or the tumour being 'medulloblastoma' do not have regard whatsoever on the diagnosis. Therefore, the attributes in our dataset every time, are said to be independent. And hence, all attributes holds equal importance (weight). For substance, the knowledge of just the symptom and origination by themselves will not be able to predict the outcome precisely. Not a single attribute is insignificant or underwritten in regard to our conclusion.
Hence, here we can state that the assumptions drawn from this Naïve bayes classifier cannot be right always in real time scenarios. Thereby we conclude saying that the independent assumptions might not always be correct, but they work with efficiency for the practice sake. Now, let's talk in brief regarding the Bayes' theorem[6].

3.2 Bayes' Theorem

It might also be known as Bayes' rule or Bayes' law, based on the prior knowledge of the event related conditions, the probability of the event is reported. Simply put, it tells the probability of the occurrence of a particular event when the probability of another event which has already occurred in given[10]. Mathematically this theorem can be expressed as follows:

$$(R|S) = \frac{(S|R)P(R)}{P(S)} \quad (1)$$

In this above equation Rand S will the events and P(R|S) will be the probability of event R occurring when event S has occurred already. Here we try to derive an event R's probability, given that the event S is always occurring (true). EventS is there by termed as evidence. P(R) is the prior probability, i.e. event's probability being noticed before indication. Here, the indication being continuously the value of an attribute of a random occurrence i.e. event S in this case. P(R|S) being the subsequent probability of S, i.e. probability of event after noticing the indication.
In above dataset,[10] the Bayes theorem is used in the following way,:

$$(b|A) = \frac{(A|b)P(b)}{P(A)} \quad (2)$$

Here, bbeing the class variable and the feature vector being A which is dependent of the size m as shown:

$$A = (a_1, a_2, \ldots, a_m)$$

As an example, we will be considering the following dataset and try to predict its outcome.

A = (Glioma, Memory Loss, Glial cells, Radiation)
b = Yes
Primarily, P(A|b) here indicates, the probability of "Occurrence" provided the conditions are "tumour is glioma", "symptom is memory loss", "origination is glial cells" and also "treatment is radiation"[4][7].

3.3 Naive assumption

Putting the below naïve assumption to the above stated rule which is independent among the features, the evidence divided into many and not just based from the reliant parts[4]. Hence, if any two events(here R and S) are being autonomous we notice that

$$P(R,S) = P(R).P(S) \quad (4)$$

Hence, the result:

$$P(b|a1, \ldots., an) = \frac{P(a1|b)P(a2|b) \ldots \ldots P(am|b)P(b)}{P(a1)P(a2) \ldots \ldots P(am)} \quad (5)$$

Which is then be expressed as:

$$P(b|a1, \ldots., am) = \frac{P(b)\prod_{i=1}^{m} P(ai|b)}{P(a1)P(a2) \ldots \ldots P(am)} \quad (6)$$

As we see here, the denominator is ruining the constant for a specific input and hence, that term can be deleted.

Then, a classifier model is created. For all the possible values of a class variable y, probability of a given set of inputs is derived and the output with extreme probability is noted. Stating it mathematically, we get [1]:

$$\hat{b} = arg(b)\prod_{i=1}^{m} P(a_i|b) \quad (7)$$

Finally, the job of deriving P(b) and $P(a_i | b)$ is left.
Do note that P(y) can also be called as class probability and $P(a_i | b)$ can be called the conditional probability. The various naive Bayes classifiers differ significantly by the choices made concerning about the distribution of $P(a_i | b)$. We apply the above formulae on our dataset to try this. For that, we first do a few pre-computations. All the other tables are classified as follows:
We try to find $P(a_i | b_j)$ for every a_i in A and b_j in b. All these derivations are shown below in their respective tables:

Occurrence Table(Table-3)

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

Treatment Table(Table-4)

Treatment	Yes	No	P(Yes)	P(No)
Radiation	3	2	3/7	2/8
Medication	2	2	2/7	2/8
Surgery	2	4	2/7	4/8
Total	7	8	100%	100%

Origin Table(Table-5)

Origination	Yes	No	P(Yes)	P(No)
Lungs	1	3	1/7	3/8
Kidney	1	1	1/7	1/8
Stomach	1	2	1/7	2/8
Glial Cells	1	1	1/7	1/8
Brain Parts	3	1	3/7	1/8
Total	7	8	100%	100%

Symptoms Table(Table-6)

Symptoms	Yes	No	P(Yes)	P(No)
Seizures	1	3	1/7	3/8
Nausea	1	4	1/7	4/8
Headache	3	0	3/7	0
Memory Loss	2	1	2/7	1/8
Total	7	8	100%	100%

Tumour Table(Table-7)

Tumour	Yes	No	P(Yes)	P(No)
Oligodendrogliomas	1	2	1/7	2/8
Secondary	2	1	2/7	1/8
Medulloblastomas	1	2	1/7	2/8
Glioma	2	1	2/7	1/8
Astrocytoma	1	2	1/7	2/8
Total	7	8	100%	100%

In the tables overhead, we intend $P(a_i \mid b_j)$ for every a_i in A and b_j in b, by ourselves. For example,probability of a tumour occurrence given the origination is stomach, that is P(origin = stomach | occurrence = Yes) = 1/7 (see table-5).The class probabilities (P(b)) are also should be found,being intended above. For substance if we consider, P(occurrence = No) will be 8/15. (see table-1)

Finally with all the assumptions we have, we are ready.

Testing this on a completely new set of features (We'll be labelling it as M):

M = (medulloblastoma, headache, glial cells, radiation)

Hence, the probability of that tumour occurring is derived as:

$$P(Yes|M) = P(Yes)P(Medulloblastoma|Yes)P(GlialCells|Yes)P(Headache|Yes)P(Radiation|Yes) \quad (9)$$

and probability of non-occurrence of tumour :

$$P(No|M) = P(No)P(Medulloblastoma|No)P(GlialCells|No)P(Headache|No)P(Radiation|No) \quad (10)$$

These are calculated with the help of the values we derived in the above tables as shown:

$$P(Yes|M) = (7/15)*(1/7)*(1/7)*(3/7)*(3/7) = 0.001749 \quad (11)$$

And

$$P(No|M) = (8/15)*(2/8)*(1/8)*(0)*(2/8) = 0 \quad (12)$$

Since P(Yes|M) > P(No|M)

So, the occurrence of the tumour is predicted as 'Yes'. For discrete data, the above discussed method is valid. In the case where the data is continuous, we might have to make some exceptions. Usually the dissimilar types of naïve Bayes classifiers vary because of the assumptions based on the distribution of $(a_i \mid b)$[1].

IV. CONCLUSION

Thereby, we conclude that various data mining algorithms can be used with ease for detecting the type of brain tumour among patients. In this paper Decision Tree and Naive Bayesian are the two classification techniques of data mining used. Both theses algorithms allow the analysis of historical data from data sets helping neurologist in predicting the type of tumour. The above prediction analysis made by the two techniques resulted that the decision tree algorithm was the quicker one and gave precise results.

REFERENCES:

1. Danda Shashank Reddy1, Chinta Naga Harshitha 2, "Brain tumor prediction using naïve Bayes' classifier and decision tree algorithms" - International Journal of Engineering & Technology , 7 (1.7) (2018) 137-141pp.1-4.
2. Kalyani A.Bhawar, Prof.Ajay S.chhajed "Brain Tumor classification using Data mining algorithms.IJESRT, November-2016, pp.239-243.
3. 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.
4. Mr. Meena, E Murali Study on Various Machine Learning Algorithms for Brain Tumor Detection.Ijpam, Volume-117 No.8 2017, pp.139-143
5. V.Vani, M.Kalaiselvi Geetha Automatic Tumor Classification of Brain MRI Images .IJCSC, Vol-4, Issue-10, pp.144-151
6. 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.
7. Kearns M. and Mansour Y. On the boosting ability of topdown decision treelearning algorithms.Journal of Computer and Systems Sciences, 58(1): 109-128, 1999.
8. 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
9. Subrata Kumar Mandal,"Performance Analysis Of Data Mining Algorithms For Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree"- IJECS Volume 6 Issue 2 Feb., 2017 Page No.20388-20391.
10. Leni Marlina1, Muslim2 "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)" -International Journal of Engineering Trends and Technology (IJETT) – Volume 38 Number 7- August 2016.

AUTHORS PROFILE

Pranitha Bhuta, Computer Science, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.

Konda Himakireeti, Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.

Nazma Mohammad, Computer Science, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.