ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

Association Rule Mining Algorithms for Brain Tumour Detection

Meghana Nagori[1], Praful Sonarkar[2] and Shivaji Mutkule[3]

Assistant Professor (Department of CSE)[1], M.E.Student(CSE)[2,3]

Government College of Engineering, Station Road, Osmanpura, Aurangabad, Maharashtra, India

Abstract: In biomedical field there is lots of data regarding patient reports, as there is daily nearly 10,000 brain tumor patient MRS image is used to find out patient suffering from which tumor type. It is difficult to predict the tumor type with the use of metabolite value like CHO, CR, CR2, and NAA for large database. Association Rule mining is one of the fundamental research topics in data mining and knowledge discovery that identifies interesting relationships between item sets and predicted the associative and correlative behaviour for new data. So author used three association rule algorithms: Apriori Association Rule, Predictive Apriori algorithm and Filtered Associator algorithm to use best rules for finding out the tumor type. These three algorithms presented at different support and confidence level, author used weka 3.7.7 data mining tool to get best rules for finding tumor type. It was found that the entire three algorithms give the best rules to easily find out the tumor type for large database.

Keywords: **f**MRI, Association Rule mining, Apriori algorithm, Predictive Apriori, Filtered Associator, Confidence level, Support level.

I. INTRODUCTION

There is problem of large database in field of biomedical field due to large number of image store in database. As in MRS images having metabolic values store in database and authors extract the metabolite value from the MRS image and store the metabolite value instead of MRS image so author uses weka 3.7.7 version to create arff file in which CHO/NAA,CHO,CR,CR2 and NAA metabolite values stored and apply filter to convert numeric to nominal after that apply data mining algorithm like Apriori Association Rule, Predictive Apriori algorithm and Filtered Association algorithm and generate rule and analyse best rule for medical application to detect the tumor type[1][2]. Data mining deals with the discovery of hidden knowledge, unexpected patterns and new rules from large databases. Data mining is often defined as finding hidden information in a database [1]. Alternatively it has been called exploratory data analysis, data driven discovery and deductive learning. Data mining essentially provides pattern-based retrieval, in which a pattern in the data is first discovered, and then that pattern is used to present information (the pattern itself or outlier data, perhaps) to the user [1][3]. The significance of the rules enervated is dictated by the pre defined minimum level of the parameter support, and the generation of rules deemed useful is guided by the pre defined minimum level of the parameter confidence [1].

II. LITERATURE SURVEY

MR Spectroscopy provides a measure of brain chemistry. The most common nuclei that are used are H (proton), Na (sodium), and P (phosphorous). proton spectroscopy is easier to perform and provide a much higher signal to noise than either sodium and phosphorous. Proton MRS can be performed within 10-15 minutes and can be added on to conventional MR imaging protocols. It can be specially used to serially monitor biochemical changes in tumor, stroke, metabolite disorder, epilepsy, infection and neurodegenerative diseases [2][4].

The basic physical principles is resonant frequencies of nuclei are at the lower end of the electromagnetic spectrum between FM radio and radar. The resonant frequencies of proton range between about 10MHz at 0.3T to about 300MHz on a 7 T magnet. The advantages of higher field strength are higher signal to noise and better separation of the metabolite peaks.

In a proton spectrum at 1.5T, the metabolites are spread out between 63,000,000 and 64,000,000Hertz. Rather than use these large numbers, some scientist decided to express the resonant frequencies in parts per million (ppm), and he/she positioned NAA at 2.0 ppm and let the other metabolites fall into their proper position on the spectral line. Then, for unknown reason if he/she reversed the ppm scale so that it real from right to left [2][4].

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

Table 1: Observable Proton Metabolites [2]

PPM	Metabolite	Properties
0.9-1.4	Lipids	Products of brain destruction
1.3	Lactate	Product of anaerobic glycolysis
2.0	NAA	Neuronal marker
2.2-2.4	Glutamine/GABA	Neurotransmitters
3.0	CreAtine	Energy metabolism
3.2	Choline	Cell membrane marker
3.5	Myo-inositol	Glial cell marker, osmolyte harmone receptor
1.2	Ethanol	Triplet
1.48	Alanine	Present in neaningiomas
3.4 & 3.8	Glucose	Increased in diabetes
3.8	Mannitol	Rx for increased ICP

For MR imaging, the total signal from all the protons in each voxel is used to make the image. If the entire signal were used for MRS, the fat and water peaks would be huge and scaling would make the other metabolite peaks invisible. Since we aren't interested in fat and water anyway, the fat and water are eliminated. Fat is avoided by placing the voxel for MRS within the brain, away from the fat in bone marrow and scalp. Water suppression is accomplished with either a CHESS (Chemical-Shift Selective) or IR (Inversion Recovery) technique. These suppression techniques are used with a STEAM or PRESS pulse sequence acquisition. A Fourier transform is then applied to the data to separate the signal into individual frequencies. Protons in different molecules resonate at slightly different frequencies because the local electron cloud affects the magnetic field experienced by the proton [2]. The STEAM (Stimulated Echo Acquisition Mode) pulse sequence uses a 90^{o} refocusing pulse to collect the signal like a gradient echo. STEAM can achieve shorter echo times but at the expense of less signal-to-noise. The RESS (Point Resolved Spectro Scopy) sequence refocuses the spins with an 180^{o} rf pulse like a spin echo. Two other acronyms require definition. CSI (Chemical Shift Imaging) refers to multi-voxel MRS. SI (Spectroscopic Imaging) displays the data as an image with the signal intensity representing the concentration of a particular metabolite [2]. As in MR imaging, the echo time affects the information obtained with MRS. With a short TE of 30 msec, metabolites with both short and long T2 relaxation times are observed. With a long TE of 270 msec, only metabolites with a long T2 are seen, producing a spectrum with primarily NAA, creatine, and choline. One other helpful TE is 144 msec because it inverts lactate at 1.3 ppm. As a general rule, the single voxel, short TE technique is used to make the initial diagnosis, because the signal-to-noise is high and all metabolites are represented. Multi-voxel, long TE techniques are used to further characterize different regions of a mass and to assess brain parenchyma around or adjacent to the mass. Multi-voxel, long TE techniques are also used to assess response to therapy and to search for tumor recurrence [2].

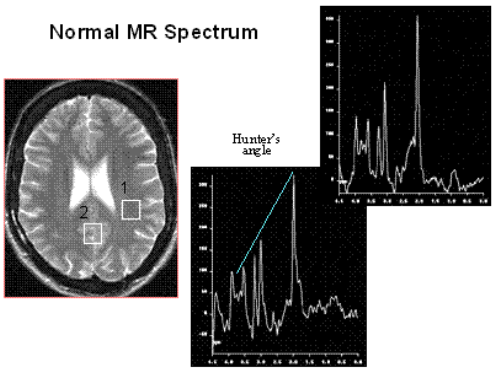

Fig. 1 Normal MR Spectrum [2]

The brain metabolites that are commonly seen on the MR spectrum are listed on the right. Each metabolite appears at a specific ppm, and each one reflects specific cellular and biochemical processes. NAA is a neuronal marker and decreases with any disease that adversely affects neuronal integrity. Creatine provides a measure of energy stores. Choline is a measure of increased cellular turnover and is elevated in tumours and inflammatory processes. The observable MR metabolites provide powerful information, but unfortunately, many notable metabolites are not represented in brain MR spectra. DNA, RNA, most proteins, enzymes, and phospholipids are missing. Some key neurotransmitters, such as acetylcholine, dopamine, and serotonin, are absent. Either their concentrations are too low, or the molecules are invisible to MRS. Normal MR Spectra obtained from gray matter and white matters are shown on the right. The predominant metabolites, displayed from right to left, are NAA, creatine, choline, and myo-inositol. The primary difference between the two spectra is that gray matter has more creatine. Hunter's angle is the line formed by the metabolites on the white matter spectrum. The common way to analyse clinical spectra is to look at metabolite ratios, namely NAA/Cr, NAA/Cho, and Cho/Cr. Normal and abnormal values are shown in the chart to the right. By including a known reference solution when acquiring the MR spectral data, absolute concentrations of metabolites can be calculated [2][4].

Table 2: Metabolite Ratios [2]

	Normal	Abnormal
NAA/Cr	2.0	<1.6
NAA/Cho	1.6	<1.2
Cho/Cr	1.2	>1.5

From NAA,Cho,Cr and Cr2 metabolite values through which author easily identify the tumour type like Benign, Malignant, Mild and Infection. MRS allows doctors and researchers to obtain biochemical information about the tissues of the human body in non-invasive way (without the need for a biopsy), where MRI only gives

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

them information about the structure of the body (the distribution of water and fat) [4].

Author's aim is to find out best rules to find out tumor type using three association rule algorithms: Apriori Association Rule, Predictive Apriori Association Rule and Filtered Associator and interpreted according to their simulation result. In general Apriori association algorithm perform better than other algorithms. Dennis P. Groth, Edward L. Robertson, "Discovering Frequent Item sets", their paper presents new techniques for focusing the discovery of frequent item sets within large, dense datasets containing highly frequent items. The existence of highly frequent items adds significantly to the cost of computing the complete set of frequent item sets [5][6].

III. PROPOSED SYSTEM DESIGN

The proposed system design takes fMRI images as an input, extract values from images and generate an arff file which is then applied to weka for association rule mining [7]. The overall system design is shown in figure 2.

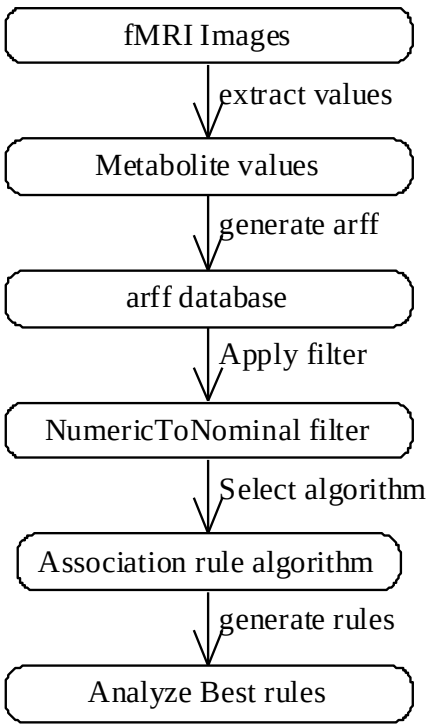

Fig. 2 Proposed System design

A. Input

The algorithmic flow of program is shown in figure 2. The input is taken in the form of MRI images. MRI images are collected from various hospitals. Our dataset consist of 159 MRI images of different patients.

These MRI images can be further classified in four distinct classes namely Benign, Mild, Malignant and Infection. The detail information is shown in Table 3.

Table 3: Number of Instances per tumor type

Tumor Type	Instances taken
Benign	45
Mild	48
Malignant	49
Infection	17
Total	159

B. Metabolite Values

MRI image contains four important metabolites namely NAA, Cho, Cr and Cr2 along with their weight. As the weights of metabolites are overlapped with graph line and images are not clear enough, so it is not possible to go for OCR technique. Few extracted weights are shown in Figure 3.

	A	B	C	D	E
1	**NAA**	**Cr**	**Cho**	**Cr2**	**Class**
2	23.63	18	17.27	11.53	Benign
3	3.93	3.03	3.77	1.97	Mild
4	12.61	9.6	9.21	6.15	Benign
5	19.7	11.22	10	4.61	Benign
6	21.71	13.16	14.59	7.59	Mild
7	9.84	6.46	14	5.46	Malignant
8	7.52	9.43	16.62	8.29	Malignant
9	4.39	4.82	5.11	2.82	Infection

Fig. 3 Sample values extracted from fMRI images

C. Arff Database

There are three ratios Cho/NAA, NAA/Cr and Cho/Cr. These are important for classifying tumor type into one of the above class. In this work, Cho/NAA is taken as one of the attribute in arff file along with above five, last attribute is taken as class attribute. ARFF file is evaluated using WEKA [11]. WEKA is suit of Machine learning algorithms.

D. NumericToNominal Filter

Generated ARFF file contains Numeric attributes. NumericToNominal filter from WEKA is used to convert numeric attributes to nominal. The filtered database is then subjected to various association rule mining algorithms [4][6].

E. Association Rule Mining Algorithms

Association rule mining algorithms are used to discover frequent patterns, association or correlation in database. This section describes Apriori, Predictive Apriori and Filtered Associator algorithm briefly.

A. Apriori Algorithm

Apriori association rule mining algorithm is used to find out frequent item sets in transaction database. Quality of association rules depends upon support and confidence [7].

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

Support: Support of an association rule $A \cup B$ is the number of transactions that contains item set $A \cup B$[1].

$$support(A \Rightarrow B) = P(A \cup B)$$

Confidence: Confidence of an association rule $A \cup B$ is the ratio of number of transactions that contains $A \cup B$ to the number of transaction that contains A [8].

$$\text{Confidence}(A \Rightarrow B) = P(B|A) = \frac{\text{support}(A \cup B)}{\text{support}(A)}$$

B. Pridictive Apriori Algorithm

This algorithm generates association rule based on predictive accuracy. Predictive accuracy is derived from support and confidence. The algorithm searches with an increasing support for the best 'n' rules. In weka, Predictive apriori has options like car, classIndex and numRules [9].

C. Filtered Associator Algorithm

Filtered Associator is a class for running arbitrary associator on data that has been passed through an arbitrary filter. Filtered associator processes the training data and test data without changing their structure [10].

IV. EXPERIMENTAL RESULTS

Apriori algorithm, predictive apriori algorithm and Filtered Associator algorithm are executed for generated database at their default configuration. All results generated are formulated in table 4. Top four rules generated by all the algorithms are same and enough to find out the tumor type of patient.

Table 4: Algorithm Output

Algorithms are used from Open source tool *i.e.* WEKA
Apriori Association Rule Algorithm
weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Minimum support: 0.1 (16 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 18 Best rules found: 1. CHO/NAA=2.21 48 ==> Result=Malignant 48 <conf:(1)> lift:(3.24) lev:(0.21) [33] conv:(33.21) 2. CHO/NAA=0.96 44 ==> Result=Mild 44 <conf:(1)> lift:(3.31) lev:(0.19) [30] conv:(30.72) 3. CHO/NAA=0.73 43 ==> Result=Benign 43 <conf:(1)> lift:(3.53) lev:(0.19) [30] conv:(30.83) 4. Result=Infection 17 ==> CHO/NAA=1.16 17 <conf:(1)> lift:(9.35) lev:(0.1) [15] conv:(15.18) 5. CHO/NAA=1.16 17 ==> Result=Infection 17 <conf:(1)> lift:(9.35) lev:(0.1) [15] conv:(15.18) 6. Result=Malignant 49 ==> CHO/NAA=2.21 48 <conf:(0.98)> lift:(3.24) lev:(0.21) [33] conv:(17.1) 7. Result=Benign 45 ==> CHO/NAA=0.73 43 <conf:(0.96)> lift:(3.53) lev:(0.19) [30] conv:(10.94) 8. Result=Mild 48 ==> CHO/NAA=0.96 44 <conf:(0.92)> lift:(3.31) lev:(0.19) [30] conv:(6.94)
Predictive Apriori Association Rule Algorithm
weka.associations.PredictiveApriori -N 20 -c -1 Best rules found: 1. CHO/NAA=2.21 48 ==> Result=Malignant 48 acc:(0.99298) 2. CHO/NAA=0.96 44 ==> Result=Mild 44 acc:(0.99262) 3. CHO/NAA=0.73 43 ==> Result=Benign 43 acc:(0.99251) 4. CHO/NAA=1.16 17 ==> Result=Infection 17 acc:(0.98239) 5. Result=Infection 17 ==> CHO/NAA=1.16 17 acc:(0.98239) 6. Result=Malignant 49 ==> CHO/NAA=2.21 48 acc:(0.982) 7. Result=Benign 45 ==> CHO/NAA=0.73 43 acc:(0.94379) 8. CHO=18.12 4 ==> CHO/NAA=2.21 NAA=8.2 4 acc:(0.90649) 9. CHO=18.12 4 ==> CHO/NAA=2.21 CR2=9.04 4 acc:(0.90649) 10. Result=Mild 48 ==> CHO/NAA=0.96 44 acc:(0.90039) 11. CHO=6.62 3 ==> CHO/NAA=0.73 NAA=9.06 3 acc:(0.87761) 12. CHO=6.62 3 ==> CHO/NAA=0.73 CR2=4.42 3 acc:(0.87761) 13. CHO=7.48 3 ==> CHO/NAA=2.21 NAA=3.39 3 acc:(0.87761) 14. CHO=7.48 3 ==> CHO/NAA=2.21 CR2=3.73 3 acc:(0.87761) 15. CHO=8.63 3 ==> CHO/NAA=0.73 NAA=11.82 3 acc:(0.87761) 16. CHO=8.63 3 ==> CHO/NAA=0.73 CR=9 3 acc:(0.87761) 17. CHO/NAA=0.73 CHO=6.62 3 ==> CR=6.9 3 acc:(0.87761) 18. CHO=2.59 2 ==> CHO/NAA=0.73 NAA=3.55 2 acc:(0.83099) 19. CHO=2.59 2 ==> CHO/NAA=0.73 CR2=1.73 2 acc:(0.83099) 20. CHO/NAA=0.73 CHO=7.77 2 ==> CR=8.1 2 acc:(0.83099)
Filtered Associator
weka.associations.FilteredAssociator -F "weka.filters.MultiFilter –F \"weka.filters.unsupervised.attribute.ReplaceMissingValues \"" -c -1 -W weka.associations.Apriori -- -N 10 -T

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Minimum support: 0.1 (16 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Best rules found:

1. CHO/NAA=2.21 48 ==> Result=Malignant 48 <conf:(1)> lift:(3.24) lev:(0.21) [33] conv:(33.21)
2. CHO/NAA=0.96 44 ==> Result=Mild 44 <conf:(1)> lift:(3.31) lev:(0.19) [30] conv:(30.72)
3. CHO/NAA=0.73 43 ==> Result=Benign 43 <conf:(1)> lift:(3.53) lev:(0.19) [30] conv:(30.83)
4. Result=Infection 17 ==> CHO/NAA=1.16 17 <conf:(1)> lift:(9.35) lev:(0.1) [15] conv:(15.18)
5. CHO/NAA=1.16 17 ==> Result=Infection 17 <conf:(1)> lift:(9.35) lev:(0.1) [15] conv:(15.18)
6. Result=Malignant 49 ==> CHO/NAA=2.21 48 <conf:(0.98)> lift:(3.24) lev:(0.21) [33] conv:(17.1)
7. Result=Benign 45 ==> CHO/NAA=0.73 43 <conf:(0.96)> lift:(3.53) lev:(0.19) [30] conv:(10.94)
8. Result=Mild 48 ==> CHO/NAA=0.96 44 <conf:(0.92)> lift:(3.31) lev:(0.19) [30] conv:(6.94)

V. CONCLUSIONS

Association rule algorithms are applied to generate rules for brain tumor detection. Experimental results show that the algorithms *i.e.* Apriori, Predictive Apriori and Filtered Associator performed well. Top eight rules generated by the algorithms are identical and important. It rules. It will help doctors to classify the patients to respective tumor type.

REFERENCES

[3] Rajeev Kumar and Arvind Kalia(Jan 2013), " A Comparative Study of Association Rule Algorithms for Investment in Related Sector of Stock Market", International Journal of Computer Application, Volume 62, No.10, pp. 32-37.
[4] John R Hesselink, "Fundamentals of MR Spectroscopy", http://spinwarp.ucsd.edu/neuroweb/Text/mrs.html
[5] Sunita B. Aher and Lobo L.M.R.J(Feb 2012). "A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning", International Journal of Computer Applications, Volume 39, No.1, pp. 48-52.
[6] M. Rudin(1992), "In vivo magnetic resonance spectroscopy", Berlin: Springer-Verlag, ISBN 3-540-55029-1.
[7] R. Agrawal, T. Imielinski, and A. Swami (1993), "Mining Associations between sets of items in Massive Databases". In proc. Of the ACM-SIGMOD 1993 international conference on Management of Data, Washington D.c USA ,pages 207-216.
[8] Ian H. Witten and E. Frank (2005), "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco.
[9] A. M. Khattak, A. M. Khan, Sungyoung Lee and Young-Koo(March 2010), "Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka", International Journal of Database Theory and Application, Vol. 3, No. 1 pp. 13-22.
[10] Mohammed M. Mazid, A.B.M. Shawkat Ali and Kevin S. Tickle(2008), "Finding a Unique Association Rule Mining Algorithm Based on Data Characteristics", 5th International Conference on Electrical and Computer Engineering ICECE.
[1] Zijian Zheng, Ron Kohavi, Llew Mason(2001), "Real world performance of association rule algorithms".
[2] He Lijun Li, Linghua Li, Xiaoniu Wang Degao(2009), "Comparison and Analysis of algorithms for Association Rules", First International Workshop on Database Technology and Applications. http://www.cs.waikato.ac.nz/~ml/weka/ accessed on 03-03-2013.

Biography

Meghana Nagori has been working as an Assistant Professor in Computer Science and Engineering department since last 10 years. Currently author is carrying out research work in the area of developing models and methods for efficient detection and classification of Brain tumour.

Praful Sonarkar received his B.E. degree in Computer Technology from Yashwantrao Chavan College of Engineering, Nagpur, India, in 2009 and pursuing M.E. degree in computer science and engineering from Government College of Engineering Aurangabad, India. His research interests include Digital Image Processing, Data Mining, Network Security, Artificial Intelligent, Machine learning and Bioinformatics.

Shivaji Mutkule received his B.E. degree in Information Technology from Vidya Pratishthan's College of Engineering, Baramati, Pune University, India, in 2011 and pursuing M.E. degree in computer science and engineering from Government College of Engineering Aurangabad, India. His research interests include Data Mining, Network Security and Bioinformatics.