European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 02, 2020

MAGNETIC RESONANCE MACHINE LEARNING METHOD FOR PREDICTING GEO GRAPHICAL LOCATION SPECIFICATION

[1]Sudhir Sharma, [2]G Shobana, [3]L Chandra Sekhar Reddy,[4]P Madhuri, [5]P Naveen

[1]Assistant Professor, Department of Computer and Science Engineering, Birla Institute of Applied Sciences, Bhimtal (Nainital), Uttarakhand – 263136.

[2]Assistant Professor, Department of Information Technology, Sri Krishna College of Engineering and Technology, Kuniamuthur, Coimbatore 641008.

[3]Assistant professor, CMR College of Engineering & Technology, Kandlakoya (v), Medchal, Hyderabad-501401.

[4]Assistance Professor, Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, Gowdavelly (V), Medchal (M), Hyderabad, Telangana-502401.

[5]Assistant Professor, Department of Computer Science and Engineering, St.Joseph's College of Engineering, OMR, Chennai - 600119.

Abstract

Graph theory is a branch of discrete mathematics that deals with the connections among entities. It has been proven to be a very beneficial and powerful mathematical tool and has a wide range of applications to handle complex problems in various domains. The aim of this work is two folds: first, to understand the basic notion of graph theory and second, to emphasis the significance of graph theory through a real-time application used as a representational form and characterization of brain connectivity network, as is machine learning for classifying groups depending on the features extracted from images. This application uses different techniques including preprocessing, correlations, features or algorithms. This paper illustrates an automatic tool to perform a standard process using images of the Magnetic Resonance Imaging (MRI) machine. The process includes pre-processing, building the graph per subject with different correlations, atlas, relevant feature extraction according to the literature, and finally providing a set of machine learning algorithms that can produce analyzable results for physicians or specialists. Further, to demonstrate the importance of graph theory, this article addresses the most common applications for graph theory in various fields.

Key terms: Magnetic resonance, Machine Learning, Preprocessing, Geo Graphical, Location Model

1. Introduction

Nowadays, the concept of the graph is becoming a mainstream of science and technology mainly because of its applications in various fields such as data mining and image

processing, networking and coding theory, clustering and scheduling, and optimization algorithms and computations. Handling an underlying problem with Graph theory to find a solution is equivalent to estimate a solution to the original real-life problem. Graph theory branches from discrete mathematics, which deals with the attributes and applications of graph [1]. It represents connections among the objects. One of the benefits of graph theory is to give a common formalism for different problems. Then, it provides graph algorithms for these problems. In all the realms, where graphs are used for modeling, the vertices or nodes represent the entities, whereas the edges represent the relationship between the entities [2].

Konignberg bridge problem is the starting point of graph theory. The solution of this famous puzzle led directly to the notion of the Eulerian graph. Euler studied the Konignberg bridge problem and found an appropriate solution in the year 1736 with the publication of Euler's solution of the Konigsberg bridge problem which is called the Eulerian graph [3]. In 1840, Mobius developed the idea of the complete graph and bipartite graph and Kuratowski proved that they are planar using recreational problems [4].

The theory of tree, (a connected graph without cycles) was developed by Kirchhoff in 1845, and he used the notions of the graph in the estimation of currents and voltages in electrical circuits. In 1852, Gutherie developed the popular four-color problem. Then in 1856, Kirkman and Hamilton investigated cycles on polyhydra and devised the idea named Hamiltonian graph by studying trips that visited certain sites exactly once [5].

2. Basic graph theoretic notations

To have a proper understanding of graph theory, it is essential to be familiar with the rudimentary terms in the graph. A graph is an ordered pair G = (V, E) containing a set V of *vertices* of *nodes* together with a set E of *edges* relating the nodes in V. Graphs are so named because they can be denoted pictorially, and this graphical representation aids us to understand many of their attributes. Nodes of a graph are denoted by dots or small circles in pictorial illustrations of graphs. An edge of a graph is represented by two nodes (e.g., n1, n2). Edges are typically denoted pictorially as either straight or curved lines joining the dots related to the appropriate nodes. It is also commonly said that nodes that share an edge are *adjacent* or *neighbors*. An edge that links two nodes is called *incident* to each of the two nodes. Edges that meet at a particular node are also known as *adjacent*. In Figure 1, the set of vertices V= {1,2,3,4,5} and the set of edges are {(1,2),(1,3),(1,4),(2,5),(3,4),(4,5)}.

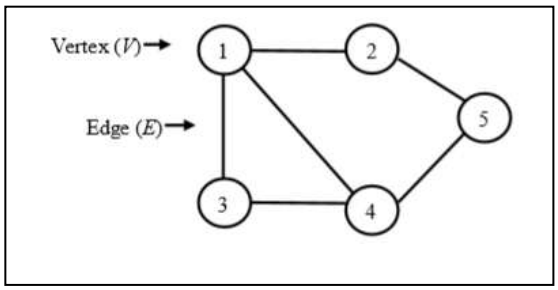

Figure 1: Graph Model

Definition 1: Simple graphs or simply graphs are graphs where no directions are designated on the edges, each edge is unique (no parallel edges), and the two ends of each edge join different nodes (i.e., there is no loop).

Definition 2: *Multigraph* is a graph that consists of more than one edge between the same two nodes (parallel edges) and edges that link a node to itself (loops) as given in Figure 2.

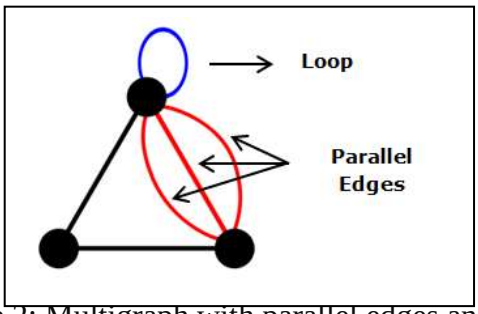

Figure 2: Multigraph with parallel edges and loop

3. System Model

Biology graphs are typically on a higher level where vertices denote amino acids and edges denote the relationship between amino acids. In computational biochemistry, there are several conditions where we need to resolve conflicts among sequences in a sample by excluding some of the sequences. Of course, exactly what constitutes a conflict must be accurately defined in the biochemical context. A conflict graph is defined where the nodes denote the sequences in the sample and there is an edge among two nodes if and only if there is a conflict among the respective sequences. The aim is to eliminate the least potential sequences that will remove all conflicts.

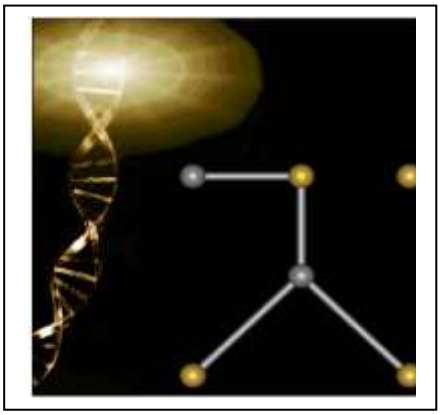

Figure 3: The DNA double helix and SNP assembly problem

Recall that given a simple graph G, a node cover C is a subset of the nodes such that every edge has at least one end in C. Therefore, the aim is to determine a minimum node cover in the conflict graph G (it is an NP-complete problem). We look at a specific example of the SNP assembly problem given in [8] and demonstrate in what way to handle this problem by means of the node cover algorithm. A Single Nucleotide Polymorphism (SNP, pronounced “snip”) is a single base mutation in DNA. It is known that SNPs are the most common source of genetic polymorphism in the human genome (about 90% of all human DNA polymorphisms).

The SNP Assembly Problem is defined as follows. An SNP assembly is a triple (S, F, R) where S = {s1, ..., sn} is a set of n SNPs, F = {f1, ..., fm} is a set of m fragments and R is

a relation R: S×F → {0, A, B} specifying whether an SNP si ∈S does not occur on a fragment fj ∈F (marked by 0) or if occurring, the non-zero value of si (A or B). Two SNPs si and sj are defined to be in conflict when there exist two fragments fk and fl such that exactly three of R(si, fk), R(si, fl), R(sj, fk), R(sj, fl) have the same non-zero value and exactly one has the opposing non-zero value. The problem is to eliminate the least possible SNPs that will remove all conflicts. The following example from is shown in Figure 4. Note that the relation R is only defined for a subset of S×F obtained from experimental values. Note, for instance, that s1 and s5 are in conflict because R (s1, f2) = B, R (s1, f5) = B, R (s5, f2) = B, R (s5, f5) = A. Again, s4 and s6 are in conflict because R (s4, f1) = A, R (s4, f3) = A, R (s6, f1) = B, R (s6, f3) = A. In the same way, all pairs of conflicting SNPs are easily calculated from the table. The conflict graph corresponding to this SNP assembly problem is shown in Figure 19.

Now, the node cover algorithm is used to determine the minimal node covers in the conflict graph. The input is the number of nodes 6, followed by the adjacency matrix of the graph given in Figure 5. The entry in row i and column j of the adjacency matrix is 1 if the nodes si and sj have an edge in the conflict graph and 0 otherwise.

0	0	0	0	1	0
0	0	0	1	0	0
0	0	0	0	0	0
0	1	0	0	1	1
1	0	0	1	0	0
0	0	0	1	0	0

Figure 5: The input for the vertex cover algorithm

The vertex cover program finds two distinct minimum vertex covers.

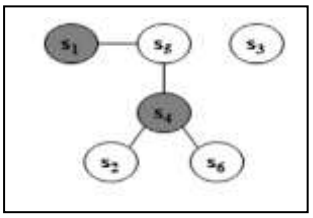

Figure 6: Minimum Vertex Cover: s1, s4

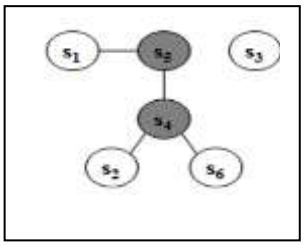

Figure 7: Minimum Vertex Cover: s4, s5

Hence, either eliminating s1, s4 or eliminating s4, s5 solves the given SNP assembly problem.

4. Map Coloring and Geo Graphical Mobile Phone Networks

Groups Special Mobile (GSM) is a mobile phone network where the geographical region of this network is decomposed into cells (i.e., hexagonal areas). Each cell has a base station that connects with mobile phones within the cell. All mobile phones connect to the GSM network by searching for cells in the neighbors. Since GSM work only in 4 dissimilar frequency ranges, it is clear by the idea of graph theory that only 4 colors can be employed to color the cellular coverage areas. These 4 different colors are employed for correct coloring of the coverage areas. Hence, the node coloring algorithm may be implemented to allocate at most 4 dissimilar frequencies for any GSM mobile phone network. Given a map drawn on the plane or on the surface of a sphere, the 4 color theorem asserts that it is always possible to color the coverage areas of a map correctly by means of maximum 4 dissimilar colors such that no two neighboring areas are given the identical color. Then, a dual graph is built by placing a node inside each coverage area of the map and links two different nodes by an edge if their corresponding areas share a whole segment of their boundaries in common. Then, the accurate coloring of the dual graph provides an appropriate coloring of the original map. Subsequently, coloring the coverage areas of a planar graph is analogous to coloring the nodes of its dual graph and vice versa. By coloring the coverage areas by means of the four-color theorem, the 4 frequencies can be allocated to the coverage areas accordingly.

The technique employed to elaborate the graph for each subject is similar to that used in other studies although the tool makes it possible to perform two types of correlation among brain sections. By using a traditional correlation or synchronization likelihood (SL), the correlation can be made by considering other aspects of the signal. Studies on graph theory often exploit various properties including the number of nodes or node degree. The proposed technique includes the three most widely used, fully described and contrasted attributes. These attributes are used in several similar studies, so they are quite relevant to characterizing the relationship among various brain sections. When implementing the technique, it automatically performs a set of 50 permutations for each subject's correlations and extracts the respective features to standardize the subject's results.

- Clustering (C): This parameter indicates the grouping of the various nodes that form the graph. A high clustering value specifies that the close relationship among brain sections activity properly but it may show a connectivity problem with distant sections.

- Path Length (L): This parameter measures the distant connections of the graph nodes from the greatest distances among the nodes found. A high path length value denotes that relationships among distant sections work correctly but that a problem may exist among sections where nodes are near each other.

- Dispersion (W): This factor estimates the relationship dispersion among various graph nodes.

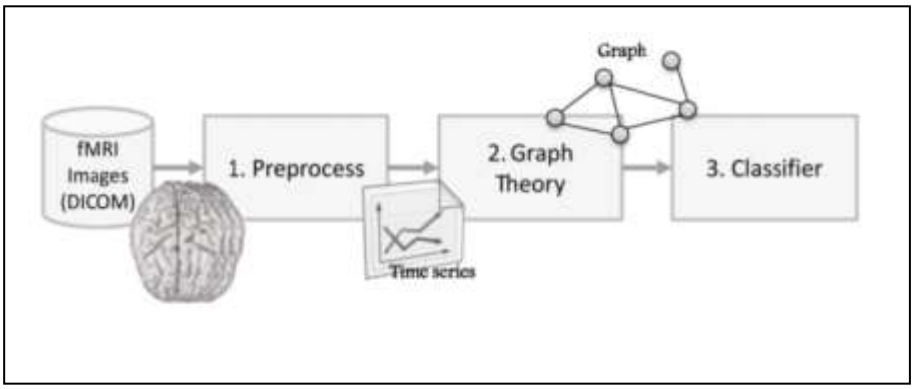

Figure 8: High-level design of the tool

1. Machine learning

The technique has a set of machine learning algorithms like Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Neural Networks (NN), K- Means, K-Nearest Neighbor and AdaBoost to categorize the groups entered in the study:

2. System design

Figure 8 shows a diagram of the high-level design of the tool. It is divided into three large blocks:

1. Preprocess: From the input of the images obtained on the MRI machine, preprocessing is performed to eliminate errors or reject imperfect images. The information from the available fMRI pictures are entered into the tool. This first step contains image alignment and motion correction, which are common problems when capturing images. This process circumvents inaccurate results or objects in the signals which makes them difficult to analyze. Once it has aligned the images, the system has a 4D image for each subject in the study. The selected atlases are employed to create the sections that will serve as masks in each of the subjects. Hence, each section of the atlas is considered to each subject's individual space, extracting a time series for each section and subject entered. Following this, the time series achieved from each image are provided.

2. Graph theory: After the time series have been extracted, various graph theory algorithms are used for extracting the features required for later classification. Correlation among all the sections for each subject are made from the time series achieved in the previous step. The system makes it possible to perform a simple correlation. A symmetric matrix with the results is achieved from the correlations. The row and columns are formed by the subjects. The system develops the corresponding graph from the matrix. The node denotes each section examined and the edges that link the connectivity defined in the matrix. Once the matrix has been estimated for each subject, the most appropriate features are extracted. In addition to the calculation for each subject, 50 permutations of each matrix are performed and the features are extracted once more so as to standardize the results.

3. Classifier: When the features have been measured, this block implements various classification algorithms, obtaining the results for each one. The input for this block consists of features extracted for each subject. And 10 subjects from one group are considered and 10 healthy subjects are considered as a practice run, leaving the rest as experimental validation. The third group is studied after the practice run, assessing which of the two main groups are almost identical. The final result is a table comparing the most commonly used algorithms, indicating the percentage of correct answers for each one.

3. Results

When the technique was implemented, a test based on images from a previous study on migraine was elaborated. To do this, a study is inserted in the technique with 90 and 116 sections of AAL regions, simple correlation and SL with the following parameters: W=1, L=1, M=6 and Pref=0.05. All values are standardized from performing 50 random permutations of the data. Table 1 shows the result of the calculations made with graph theory. For each group (CON: controls, MIG: sporadic migraine, ABU: medication abuse). The mean value and the standard deviation of the three features are examined (clustering, path length, dispersion). The results are presented for all the sections (116) and 90 sections excluded the cerebellum sections by means of AAL atlas.

Table 1 Graph theory results.

Charact.	Areas	CON		MIG		ABU	
		M(N/SL)	SD(N/SL)	M(N/SL)	SD(N/SL)	M(N/SL)	SD(N/SL)
C	90	1.075/1.069	0.024/0.028	1.08/1.069	0.012/0.026	1.076/1.057	0.013/0.02
	116	1.065/1.084	0.02/0.027	1.07/1.09	0.012/0.02	1.066/1.074	0.011/0.021
L	90	1.007/1.044	0.001/0.008	1.007/1.048	0.001/0.007	1.007/1.045	0.001/0.007
	116	1.004/1.03	0.001/0.005	1.004/1.032	0.001/0.005	1.004/1.031	0.001/0.004
W	90	0.995/0.807	0.007/0.018	0.997/0.804	0.001/0.017	0.995/0.8	0.002/0.017
	116	0.996/0.834	0.004/0.016	0.997/0.828	0.001/0.015	0.9962/0.822	0.002/0.016

From the data in Table 1, a classification with different classifiers, areas, and correlations was performed. The results are shown in Table 2. Sensitivity and Specificity results are noted. The sensitivity indicates the classifier's capability to classify disease in sick subjects, while Specificity specifies the classifier's capability to classify the absence of disease. The proposed technique can carry out the whole process, from taking fMRI images to provide final results which can be directly understood by physicians or specialists. This is an automated process, where the user only submits fMRI images and selects the correlations and the best atlas to use. Analysis of subjects suffering from migraine and drug abuse was carried out to evaluate this technique. The technique conducts a complete analysis and proposes different classifiers, some of which deliver 92.86% accuracy (NN), while others offer 87% (SVM). Other studies in other pathologies with similar machine learning algorithms had success rates as between 75% and 87.9%, so the proposed technique has revealed acceptable results. The existing differences between classifiers' results are due to many reasons including their type (supervised, unsupervised, semi-supervised) or the difference between classifiers with the same data which might attain global or local performance.

5. Conclusion

This paper explores different elements involved in graph theory including graph representations using computer systems and graph-theoretic data structures such as list structure and matrix structure. To emphasis the significance of graph theory, this paper describes a real-time application used as a representational form and characterization of brain connectivity network, as is machine learning for classifying groups depending on the features extracted from images. This application uses different techniques including preprocessing, correlations, features or algorithms. This paper illustrates an automatic tool to perform a standard process using images of the MRI machine. The process includes pre-processing, building the graph per subject with different correlations, atlas, relevant feature extraction according to the literature, and finally providing a set of machine learning algorithms that can produce analyzable results for physicians or specialists. Further, to demonstrate the importance of graph theory, this article addresses the most common applications for graph theory in various fields. This paper also presents a survey on the graph theory challenges relevant to their approaches and techniques.

References

1. Bondy J A and Murty U S R., Graph theory with applications, Elsevier science publishing, 1976.
2. Caccetta L. and Vijayan K, Applications of graph theory, Fourteenth Australasian Conference on Combinatorial Mathematics and Computing, Ars Combinatoria, Vol. 23, 1987, 21-77.
3. Gary Chartrand, Ping Zhang, A First Course in Graph Theory, Dover books on mathematics, Courier Corporation, 2012.
4. Johan M.M. van Rooij, Hans L. Bodlaender.: Exact algorithms for dominating set. Discrete Applied Mathematics, 159 (2011), 2147–2164.
5. Shariefuddin Pirzada and Ashay Dharwadker, "Applications of Graph Theory", Journal of the Korean Society for Industrial and applied Mathematics, Volume 11, No.4,2007.
6. Narasingh; Deo; "Graph theory with applications to engineering and computer science", Prentice Hall of India, 1990.
7. Alexandru T. Balaban, Journal of Chemical Information and Computer Science, 1985, 25 (3), pp 334–343.
8. G. Lancia, V. Afna, S. Istrail, L. Lippert, and R. Schwartz,SNPsProblems, Complexity and Algorithms, ESA 2002, LNCS 2161, pp. 182-193, 2001. Springer-Verlag 2001.
9. Ashay Dharwadker, The Vertex Coloring Algorithm, 2006
10. Ashay Dharwadker, " a new proof of The four colour theorem", Institute of Mathematics, Haryana, India.
11. Eric Filiol, Edouard Franc, Alessandro Gubbioli, Benoit Moquet and Guillaume Roblot, Combinatorial Optimization of Worm Propagation on an Unknown Network, Proc. World Acad. Science, Engineering and Technology, Vol 23, August 2007.
12. Daniel Marx, "Graph Coloring problems and their applications in scheduling"
13. M. Abellanas, F. Hurtado, C. Icking, R. Klein, E. Langetepe, L. Ma, B. Palop, and V. Sacristan. "The farthest color Voronoi diagram and related problems". In Abstracts 17th European Workshop Comput. Geom., pages 113–116. Freie University¨at Berlin, 2001.
14. Manjula V (2010) Electrical networks by graph theory. Int J of Math and Eng IEEE 149: 1368 -1373.
15. M. A. Rajan, M. Girish Chandra, Lokanatha C. Reddy, Prakash Hiremath "Concepts of graph theory relevant to Adhoc Networks", Int. Journal of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 18419844 Vol. III 2008.
16. C. Stam, W. De Haan, A. Daffertshofer, B. Jones, I. Manshanden, A.V.C. vanWalsum, T. Montez, J. Verbunt, J. De Munck, B. Van Dijk et al., Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease, Brain 132 (2009), 213–224.
17. J. Wang, X. Zuo and Y. He, Graph-based network analysis of resting-state functional MRI, Frontiers in Systems Neuroscience 4 (2010), 1–14.
18. M.P. van den Heuvel and H.E. Hulshoff Pol, Exploring the brain network: a review on resting-state fMRI functional connectivity, European Neuropsychopharmacology 20 (2010), 519–534.
19. L. He, D. Hu, M. Wan and Y. Wen, Measuring temporal dynamics of resting-state fMRI data, Bio-Medical Materials and Engineering 24 (2014), 939–945.