International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

An Algorithmic Approach for Alzheimer's Disease detection from Non-Image Data

C R Aditya†* and M B Sanjay Pande†

†Department of CSE, VVIET, Mysuru, Karnataka–570028, India

Accepted 15 May 2016, Available online 18 May 2016, **Vol.6, No.3 (June 2016)**

Abstract

In this paper an exploratory data analysis model is proposed to create a suitable reference knowledge base from Alzheimer's disease dataset. The knowledge of the reference base is expressed in terms of zones with each zone carrying a weightage factor. The learnt knowledge is used to quantify the similarity of a test sample with respect to the demented class. Evaluation of the model on OASIS longitudinal database of Alzheimer subjects shows that the designed model successfully explores the data set for useful information and assigns test samples to either non-demented or demented class with non-overlapping and measurable similarity indices.

Keywords: *Alzheimer's Disease, Dementia, Knowledge Discovery, Similarity Measure, Affiliation Analysis.*

1. Introduction

Alzheimer 's disease (AD) is the most common type of dementia disease involving degeneration of the brain which is irreversible and gradually ends up with the complete brain failure. According to the statistics of Alzheimer's Association, AD accounts for 60 to 80 percent of the dementia cases (Alz.org, 2016). It is expected to double by 2030 and triple by 2050 as projected by world health organisation. In 2006, there were 26.6 million sufferers worldwide, and AD is predicted to affect 1 in 85 people globally by 2050, and at least 43% of prevalent cases need a high level of care (Brookmeyer *et al*, 2007; Alzheimer.ca, 2013). Aging and other factors increase the possibility of neuron degeneration and can lead to AD. Many studies are made and models are created to analyse and detect AD from MRI images (Mahanand *et al*, 2011; Zhang *et al*, 2014; Babu *et al*, 2012). Techniques such as neural networks, support vector machines, decision tree classifiers are successfully applied on MRI images to find the region of interest responsible for causing AD (Mahanand *et al*, 2012; Kubota *et al*, 2006; Pennanen *et al*, 2004). But only few studies have explored the possibility of AD detection from non-image data (Ertek *et al*, 2014). In medical data analysis such as detecting AD, it is important that a standard measurable scale is formulated to compare the severity of the disease.

Computational models provide a better means of modelling complex systems (such as the nervous system, neuro-disorders, etc.). To arrive at a good model for analysis, we need to design the knowledge base (KB) and ensure accuracy in the classification of data samples. Creation of a reference KB implies consolidating a huge database of control/training set into knowledge parameters, which provide meaningful and useful comprehension for later analysis. It is required to affiliate whether the sample is normal or abnormal on a measurable scale. In this study we explore the possibility of analysing AD subjects from non-image data using a parametric model and quantify their abnormality using measurable similarity indices.

The main theme of this research paper is to devise a computational approach for critical analysis of an input (a test sample). Two phases are involved in the suggested computational model: (i) Learning Phase (ii) Recognition Phase. Learning phase involves building up a reference KB using known healthy and unhealthy samples. In the recognition phase a test sample is contrasted with the reference KB to decide the label (Normal/Abnormal), and further the degree of belongingness (affinity) of the sample with respect to abnormal class.

2. Materials and Methods

2.1 Data Set

The proposed model is evaluated on data obtained from the Open Access Series of Imaging Studies (Marcus *et al*, 2010). The dataset contains both men and women subjects who are all right-handed. The data also includes the education level and socio-economic status of the subjects. Moreover, some other medical statistics exist in the dataset, including intracranial volumes and brain volumes of the subjects. The two

*Corresponding author: **C R Aditya**

groups of subjects are demented and non-demented in which the patient has the AD or not, respectively. Patients that develop the AD during the tests are grouped as converted. Clinical Dementia Rating (CDR) can only take values 0, 0.5, 1 and 2. CDR being equal to 0 corresponds to non-demented subject CDR being equal to 0.5, 1, 2 corresponds to very mild dementia, moderate dementia and severe dementia respectively.

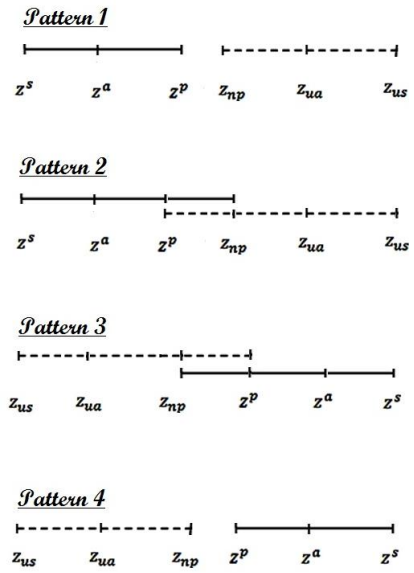

Fig.1 Generic distribution patterns of sample values. Solid line indicates normal sample distribution and dotted line indicates abnormal sample distribution

2.2 Knowledge-Base Creation

The assignment of variables into two groups must always be motivated by the nature of the response variables and never by an inspection of data. Majority of physical and mental traits tend to be distributed as to approximate normal distribution stretching from to + and covering the unit area interpreted as probability of occurrence of the universe of traits or events which describes, the normal distribution as an ideal. But to make a computational model more generic, it is required to design the model independently from any specific distribution type. In this work we propose four generic distribution patterns that are likely to occur in real world, and which can be used to analyse the distribution pattern of any samples. These generic distribution patterns are decomposed several mutually exclusive zones and weightage factor is assigned to each zone. Later, affiliation analysis of a test sample is performed by a suitable distance measure. The discrimination efficacy of the new samples either as healthy or not, depends on the learning done on the known healthy and unhealthy data set. The strength of knowledge derived mainly depends on distribution pattern of each feature.

We wish to divide the distribution space based on the observations of training data set into m mutually exclusive and exhaustive regions R1.....Rm. The distribution of normal and abnormal samples in medical data analysis exhibit one of the four patterns shown in figure 1. The distribution patterns shown are not based on probability distribution function, but merely on the range of values the samples can take. Each of the distribution pattern is split into five zones: Safe [z^s - z^a], Acceptable[z^a - z^p], Permissible[z^p - z_{np}], Unacceptable[z_{np} - z_{ua}] and Unsafe[z_{ua} - z_{us}]. If a test sample lies in the safe zone, then it can be declared to be safely healthy. Similarly a test sample can be declared to be acceptable healthy or permissibly healthy . If a test sample lies beyond the permissible zone then it is to be considered unhealthy. To express these qualitative discriminations on a number scale, suitable weightage factors are attributed to different zones. Weights 0.01 , 0.1, 1, 10, 100 are chosen for the zones Safe, Acceptable, Permissible, Unacceptable and Unsafe respectively. The higher weights are assigned for zones that represent the higher abnormality regions. The knowledge base is expressed in terms of zones for each feature in the sample space and their respective weights.

The algorithm for creating the knowledge base are as follows:

Input: n-dimensional feature values [f1 , f2 , . . . fi ,. . . fn,] for *m* AD and *m* ND samples.
Output: KB containing information of feature wise zones.
For each feature *fi* in the sample space perform the following,
Step 1: Calculate min, max and mean ND (min_{ni}, max_{ni}, μ_{ni})and AD (min_{ai}, max_{ai}, μ_{ai}) samples.
Step 2: Fit a distribution pattern using the above details (min, max and mean).
Step 3: Decompose the area under the distribution into different zones (Safe, Acceptable, Permissible, Unacceptable and Unsafe zones).
Step 4: Formulate the KB for the i-th feature in terms of zone-ranges: z^s , z^a, z^p, z_{np}, z_{ua}, z_{us}.

2.2 Affiliation Analysis

The affiliation algorithm used in our model estimates the degree of test sample being away from the ND samples on a measurable scale. When a test sample is presented for affiliation, the feature values are initially compared with the knowledge base to determine the zone in which each feature value lies. Respective zone weights are obtained from the KB and the overall affiliation is done through the similarity index computed by summing up the individual feature distances. The distance measure used to calculate the feature distance is derived from Euclidian distance.

The Euclidean distance between two points (p,q) and (r,s) is given by

$$D_i = \sum \sqrt{(p - r)^2 + (q - r)^2} \quad (1)$$

We use the idea behind Euclidean distance to find the feature wise distance of a sample with respect to the mean of the feature in consideration. Let μi denote the mean value for the i[th] feature in a training data set containing equal number of ND and AD samples. Given the mean value μ_i along with the feature value f_i of a test sample, the Feature Distance is calculated using equation-2. The term ($\mu_i - f_i$) denotes the farness / closeness of a sample for a selected feature with respect to the mean value. The Affiliation Distance for a sample is the sum of all the individual feature distances, or in other words it is the final Feature Distance after processing all the feature values (i = n − 1 in equation-2). W_i is the zone weight matched for i[th] feature.

$$FD_{i+1} = \sum \sqrt{(FD_i)^2 + w_i * (\mu_{i+1} - f_{i+1})^2} \quad (2)$$

The algorithm for affiliation analysis is as follows:

Input:
1. KB of zone details of all n-d feature space
2. Weight age factors for different zones
3. Test sample with n-dimensional feature values [f1 , f2 , . . . fi ,. . . fn,]
Output: AD-Affiliation distance for the test sample.

Step 1: For a feature "'i"' match the distribution pattern from the KB.
Step 2: Check in which zone of the matched distribution pattern, does the feature value f_i of the test sample lie. Decide the weightage factor w_i accordingly.
Step 3: Calculate the Affiliation-Distance by using equation-2. Multiply w_i to the calculated distance.
Repeat the till all the features are processed.

3. Results

Experimentation was carried out on the Alzheimer's data obtained from OASIS. From the data set selected, the samples with missing values were discarded and only the samples with complete data entries were input to the experimentation. *Group* attribute is considered as target label for all experiments. The samples for training and testing phase were input without considering the subject-ID,MRI-ID and visit attributes of the samples. Two different samples of a same subject was considered to be different samples for training the system. In total the samples were assessed with respect to the eight attributes - *Age, Education, SES, MMSE, CDR, eTIV, nWBV, ASF.* Since the subjects in the data set are only right handed, the *Hand* attribute was not considered for experimentation, as it would not affect the knowledge-base creation phase nor affiliation analysis of a test sample.

50 samples of ND and 50 samples of AD were chosen randomly for the knowledge base creation. Table-1 provides the range of distribution zones for each feature of the subjects used for experimentation. Table-2 summarizes the demented affiliation distances obtained for test samples which comprised of randomly chosen ND and AD samples.

The affiliation of a test sample is assessed across each dimension by comparing a feature value with the distribution zones created. Analysis of data from the zones created infers that demented subjects have relatively lesser social economic status, normalized whole brain volume and MMSE index than non-demented subjects. The range of affiliation distances obtained for ND and AD samples have overlapping affiliation indices. As the abnormality of the sample increases, the affiliation distance measured quickly rises towards significantly larger value as compared to normal samples. Table-3 shows the refined affiliation distances for quantifying AD samples, after discarding the overlapping affiliation indices from the result obtained. A total 13 of 100 affiliation indices were discarded (classification accuracy = 87%).

Conclusion

An exploratory data analysis based model is proposed to carry out affiliation analysis by designing a generic distribution-zonalization model. The knowledge of the reference base is expressed in terms in terms of zones with each zone carrying a weightage factor. The designed computational model is able to match the test sample with KB, incorporating the weightage factor assigned based on the location a test sample falls, thus understanding the closeness/farness of the test sample with respect to the abnormal class. The non-demented and demented subjects are affiliated by a range of separable distance values through a novel distance measure.

The future work will extend the proposed model to explore inter-feature relationships in a data set using more sophisticated algorithm. The model will be enhanced to handle different levels of severity in demented class - mild, moderate, high.

Table 1 Range of distribution zones for 50 ND and 50 AD samples

	AGE	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Zs	60	8	1	15	0	1143	0.65	1.01
Za	63	10	1	21	0	1171.5	0.66	1.03
Zp	67	12	1	27	0	1200	0.68	1.05
Znp	97	18	4	30	0.5	1732	0.77	1.6
Zua	97	20	4.5	30	1.25	1698	0.79	1.53
Zus	98	23	5	30	2	1665	0.82	1.49

Table 2 Range of AD-affiliation distance obtained for 100 test samples containing 50 samples of both ND and AD samples

	AD affiliation distance (Female)	
Sample	Min	Max
Non-Demented	0.009	262.46
Demented	27.37	896835633

Table 3 Refined AD-affiliation range

	AD affiliation distance (Female)	
Sample	Min	Max
Non-Demented	0.009	262.46
Demented	415.44	896835633

References

Alzheimer.ca, (2013), World health organisation report. URL: http://www.alzheimer.ca/en/Get-involved/Raise-your-voice/WHO-report-dementia-2012. [Online; accessed 24-January-2013].

Alz.org, (2016), Alzheimer's and dementia. URL: http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp. [Online; accessed 2-February-2016].

Babu, G.S., Suresh, S., Mahanand, B., (2012), Alzheimer's disease detection using a projection based learning meta-cognitive rbf network, *The 2012 International Joint Conference on Neural Networks (IJCNN)*, *IEEE*. pp.1–8.

Brookmeyer, R., Johnson, E., Ziegler-Grahamm, K., Arrighi, H.M., (2007), Forecasting the global prevalence and burden of alzheimers disease. *Alzheimer's & Dementia* 3, S168.

Ertek, G., Tokdil, B., G¨unaydın, ˙I., (2014). Risk factors and identifiers for alzheimers disease: A data mining analysis, *Advances in Data Mining : Applications and Theoretical Aspects. Springer*, pp. 1–11.

Kubota, T., Ushijima, Y., Nishimura, T., (2006). A region-of-interest (roi) template for three-dimensional stereotactic surface projection (3d-ssp) images: Initial application to analysis of alzheimer disease and mild cognitive impairment, *International Congress Series, Elsevier.* pp. 128–134.

Mahanand, B., Suresh, S., Sundararajan, N., Kumar, M.A., (2011). Alzheimer's disease detection using a self-adaptive resource allocation network classifier, *The 2011 International Joint Conference on Neural Networks (IJCNN)*, *IEEE*. pp. 1930–1934.

Mahanand, B., Suresh, S., Sundararajan, N., Kumar, M.A., (2012). Identification of brain regions responsible for alzheimers disease using a self-adaptive resource allocation network. *Neural Networks* 32, 313–322.

Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L., (2010). Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. *Journal of cognitive neuroscience* 22, 2677–2684.

Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., H¨anninen, T., Laakso, M.P., Hallikainen, M., Vanhanen, M., Nissinen, A., Helkala, E.L., *et al*., (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of aging* 25, 303–310.

Zhang, Y., Wang, S., Dong, Z., (2014). Classification of alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Progress in Electromagnetics Research* 144, 171-184.