

SPARSE REPRESENTATION BASED CLASSIFICATION OF MR IMAGES OF BRAIN FOR ALZHEIMER'S DISEASE DIAGNOSIS

Vinith Rejathalal¹, Srinu Babu Mandalapu², V K Govindan³

¹Ph.D, Department of Computer Science and Engineering, NIT Calicut, Kerala, India

²M.Tech, Department of Computer Science and Engineering, NIT Calicut, Kerala, India

³Ph.D, Department of Computer Science and Engineering, NIT Calicut, Kerala, India

Abstract

High-dimensional pattern classification methods, like support vector machines (SVM), have been widely used for analysis of structural and functional brain images to assist the diagnosis of Alzheimer's disease (AD). However, due to noise and small sample size of neuroimaging data, it is challenging to train a global classifier that can be robust enough to achieve good classification performance. In this paper, we evaluate the performance of the SVM and Dictionary learning method such as K-SVD by using a local patch based method. Dictionary Learning involves finding a basis set (dictionary) that best represents the data and finding a sparse representation in terms of the dictionary. We evaluated our method on 165 subjects (including 55 AD patients, 54 MCI and 56 normal controls) from Alzheimer's disease Neuroimaging Initiative (ADNI) database using MR images. The K-SVD method gives better results than SVM by using local patch based method. By using K-SVD getting very promising performance when compared to SVM method for AD/MCI classification using MR images.

Keywords: Sparse representation; Feature extraction; dictionary learning; K-SVD; local patch

-----***-----

1. INTRODUCTION

Alzheimer's disease (AD) is the most frequent neurodegenerative dementia and a growing health problem. Early and accurate diagnosis of Alzheimer's Disease (AD) is not only challenging, but is crucial in the perspective of future treatment. There are basically 3 stages in Alzheimer disease. Those are cognitive normal (CN), mild cognitive impairment (MCI) and Alzheimer disease (AD) [1]. The key to early diagnosis and treatment is the recognition of mild cognitive impairment. Patients suffering from AD at a prodromal stage are, mostly, clinically classified as amnesic mild cognitive impairment (MCI) but not all patients with amnesic MCI will develop AD. Distinguishing mild cognitive impairment from cognitive normalcy by clinical means is a challenging task. The onset of cognitive impairment is recognizable as structural abnormality in the brain even before the subject begins to show clinical symptoms. This makes brain imaging a potential tool in early diagnosis of Alzheimer's Disease. The modality we used here is T1 weighted MRI. Advantages of MRI over other imaging modalities are its high spatial resolution and excellent discrimination of soft tissues [2].

Advances in statistical learning with the development of new machine learning algorithms capable of dealing with high dimensional data, such as support vector machine (SVM), enable the development of new diagnostic tools based on T1-weighted MRI. Recently, several approaches have been proposed to automatically classify patients with AD and/or

MCI from anatomical MRI. These approaches could have the potential to assist in the early diagnosis of AD. These approaches can roughly be grouped into three different categories, depending on the type of features extracted from the MRI (voxel-based, vertex-based or ROI-based). In the first category, the features are defined at the level of the MRI voxel. Specifically, the features are the probability of the different tissue classes (grey matter, white matter and cerebrospinal fluid) in a given voxel directly classified grey matter features with an SVM [3]. All other methods first reduce the dimensionality of the feature space using different types of features extraction and/or selection methods used smoothing, voxel-down sampling, and then a feature selection step. Another solution is to group voxels into anatomical regions through the registration with a labeled atlas. In the second category, the features are defined at the vertex-level on the cortical surface. The methods of the third category include only the hippocampus [4]. Their approach is based on the analysis of the volume and/or shape of the hippocampus. The goal of this work is to investigate the efficacy of sparse dictionary learning in addressing the problem of classification of patients with AD based on anatomical MRI, using the same study population. To that purpose, we used the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We evaluated the performance of sparse learning against SVM which is one of the most successful data classification tools [5]. Voxel based approach was used for feature selection.

The paper is organized into 5 sections. Section 1 presents an introductory discussion of the problem. Section 2 introduces Sparse Dictionary Learning. Section 3 is about Data and Methods used such as patch extraction and sparse method of classification. Section 4 presents the results and discussion and finally Section 5 concludes the paper.

2. SPARSE REPRESENTATION

Recent years have seen a surge in interest in the sparse representation. A collection of data vectors are represented using an over-complete dictionary, $D \in \mathbb{R}^{n \times K}$ where 'n' is the dimension of the data-space and 'K' is an integer greater than n. Assuming a linear model of data generation the data vectors are represented as a linear combination of the dictionary-atoms [6]. Since the number of dictionary-atoms is greater than the dimension of the data-space it is possible to have an exact representation of any vector in the data-space. In sparse representation the accuracy of the linear combination is coupled with a sparsity factor. This results in generating representation that is accurate and sparse at the same time. In simple mathematical terms sparse representation problem can be expressed in either of the following ways.

$$\min \|x\|_0 \quad \text{s.t. } y=Dx \quad (1)$$

or

$$\|x\|_0 \quad \text{s.t. } \|y-Dx\|_2^2 \quad (2)$$

Where $\|a\|_p$ is the l_p norm defined as $(\sum |a_i|^p)^{1/p}$ $p > 0$

and a_0 is the number of non-zero elements in 'a'. Extraction of the sparsest representation is a hard problem that has been extensively studied. In sparse-coding [7] the dictionary D is assumed to be known a priori and fixed. The drawback of this approach is that the dictionary might not be suitable to capture the intricate characteristics of the data. This innate deficiency of fixed dictionary lead to the development of algorithms that deduce prototype vectors from the available data samples itself. Section II A introduces the sparse-coding stage based on a fixed dictionary. Section II B discusses the update on the dictionary-atoms to better fit the data.

2.1 Sparse Coding

Sparse-coding is the process of computing the representation x of a data vector y given the dictionary. Exact solution of the problems 1 or 2 is NP hard and hence approximate solutions are sought for. In this paper we discuss the simplest of the approaches namely the Orthogonal Matching Pursuit (OMP) which is a generalization of the clustering problem. Clustering can be interpreted as identifying the closest vector from a group of prototype vectors and assigning a label to the data vector according to the prototype vector it is most closely associated. Clustering provides a compressed representation

of the data where each data sample is approximated by its closest prototype vector. Hence, it can be construed as a special case of sparse-coding where for each data the representation is composed of a single non-zero component [8]. In our work, the sparse-coding phase is addressed using OMP algorithm. The pseudo-code listed taken from [9] below gives the OMP algorithm. OMP is an iterative procedure where in the i^{th} iteration the i^{th} closest dictionary atom is chosen and the projection of the data-vector on the selected dictionary-atom is added to the representation vector. This is iterated L times where L is the required l_0 norm of the sparse-code. The optimization problem that OMP addresses is:

$$\min \|y-Dx\|^2 \quad \text{s.t. } \|x\|_0=L \quad (3)$$

Algorithm 1: Orthogonal Matching Pursuit

Input: signal b, and matrix A.

stopping criterion e.g. until a level of accuracy is reached.

Output: Approximation vector c.

Algorithm.

1. Start by setting the residual $r_0 = b$, the time $t = 0$ and index set $V_0 = \emptyset$;
2. Let $v_t = i$; where a_i gives the solution of $\max \langle r_t; a_k \rangle$
3. where a_k are the row vectors of A.
4. Update the set V_t with v_t : $V_t = V_{t-1} \cup \{v_t\}$
5. Solve the least squares problem

$$\min_{c \in \mathbb{R}^{V_t}} \|b - \sum_{j=1}^t c(v_j) a_{v_j}\|_2$$

6. Calculate the new residual using c

$$r_t = r_{t-1} - \sum_{j=1}^t c(v_j) a_{v_j}$$

7. Set $t = t + 1$
8. Check stopping criterion if the criterion has not been satisfied then return to step 2

2.2 Sparse Dictionary Learning

Once a representation of the data with respect to an initially assumed dictionary is computed the next step is to update the dictionary [10]. Each dictionary atom is updated to make it closer to the subset of data-vectors for which it has a non-zero contribution in the sparse representation. Denoting by Y the collection of data-vectors and by X the sparse-codes of the data-vectors obtained in the previous stage the error associated is defined as:

$$E=Y-DX \quad (4)$$

This can be expanded using the dictionary atoms:

$$E = Y - \sum d_i x_i^i \quad (5)$$

Separating the term containing the relevant dictionary atom we get an equivalent expression

$$E = E_k - d_k x_T^k \quad (6)$$

Where $E_k = y - \sum_{i \neq k} d_i x_i^T$.

Dictionary update in KSVD algorithm [11] minimizes the error by minimizing the difference between E_k and $d_k x_T^k$. That is it finds nearest rank-1 approximation to E_k and updates d_k . The best approximation is obtained by the using the vectors corresponding to the largest singular value of E_k . The algorithms involving the two stages are given below [12].

Algorithm 2 KSVD

1. Input: NXM Matrix of data vectors Y, N is the dimension, M is the no. of vectors and K is the number of dictionary-atoms.
2. Output: Sparse Representation Y, Dictionary D.
3. Init: D = random NKK dictionary.
4. Compute X = OMP(Y,D) //compute the representation of Y w.r.t D
5. For each atom d_k
 - a. $E_k = Y - \sum_{i \neq k} d_i x_i^T$
 - b. $[U_k, \Lambda_k, V_k] = \text{SVD}(E_k)$
 - c. $d_k = U_k(:,1)$
6. If $\|E\|_2 < \Theta$ go to step 4.

3. DATA AND PROPOSED METHODS

The objective of the work as stated earlier is to investigate the potential of sparse learning in AD diagnosis. We choose to use the standard bench mark data ADNI presented in the Table 1 for our study.

Table-1: Standard bench mark data ADNI

Diagnosis	Number	Age	Gender(M/F)
AD	55	76±7	35/20
MCI	54	74±7	33/21
CN	56	72±8	33/23

The dimensions of the brain volume are 91x109x91. Since this work is preliminary investigation to test our intuition about the possible use of sparse learning methods as a diagnostic tool for AD, we decided on using a single slice containing enough cortical region to extract voxel features. 55th slice was used for the study. Below Fig 1 shows the relevant slice of brain and brain affiliated with AD and MCI.



Fig-1: Normal, AD, MCI

The volumes were preprocessed before the slice and the voxels were selected. The preprocessing steps are basically done in 3 stages- Brain extraction, image registration and segmentation respectively.

Preprocessing was done using FSL toolbox

<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>. MRI image contain some unnecessary portions like skull and noise. Using the Brain Extraction we will get only the portion of brain excluding the skull. Image registration is used to register all other images to a standard template image. MNI152 is used a standard template image. The brain extracted, co-registered image is segmented into GM, WM and CSF using FAST segmentation tool box in FSL. We chose grey matter for the classification.

Since the slice chosen contains a large number of cortical voxels, its direct vectorization results in high dimensionality. To address this problem we chose patches uniformly distributed over the cortical surface. These patches are vectorized and concatenated to obtain a vector representing the slice. Since the patch distribution is random and uniform, when taken over a number of images they collectively contain enough features for the algorithms to distinguish the underlying classes. The algorithm of feature selection is described below.

3.1 Patch Extraction Algorithm

After extraction of the grey matter, from the MR images, it is required to select the patches. For simplicity, we uniformly select the patches of fixed size(pxp) from the MR images. We select the significant j number of patches from each image and used the data matrix for the training of SVM and K-SVD.

1. Select the particular slice having the large number of cortical voxels.
2. Select a large number N
3. For $i = 1$ to N // i is some large number
 - a. Select two pixel positions x, y using uniform distribution function.
 - b. Form a matrix S by taking x, y as the origin (pxp).
 - c. $S = \text{slice}\left(x - \left\lfloor \frac{p}{2} \right\rfloor : x + \left\lfloor \frac{p}{2} \right\rfloor, z - \left\lfloor \frac{p}{2} \right\rfloor : z + \left\lfloor \frac{p}{2} \right\rfloor\right)$
 - d. If number of non zero values were less than n then go to 3.

- e. Otherwise, make the S as a vector. Then concatenate S into a matrix $B = \text{vertcat}(B, S)$;
- f. $\text{count} = \text{count} + 1$
- g. if $\text{count} = j$
 - i. break;

The vectors obtained as detailed above were used as input data for both SVM and Sparse dictionary method.

3.2 Sparse Method of Classification

The patch extraction algorithm extracts N patches of size $P \times P$ from a slice and concatenates them into an $NP^2 \times 1$ vector. The vectors extracted so form the training data for each class. Denote it by $Y_i^T \in R^{NP^2 \times n_i}$ the training data for class i , $i \in \{\text{AD}, \text{MCI}, \text{CN}\}$, where n_i is the number of training images in class i . The training vectors are used by KSVD algorithm to extricate the codebook vector dictionary for each class i . Training stage in Sparse representation based classification is the deduction of the codebook dictionaries from the training data. Training stage yields C dictionaries $D_i \in R^{NP^2 \times K}$, where C (=3) is the number of classes to be recognized. The classification algorithm using the class dictionaries is based on the intuition that a general data vector $Y \in R^{NP^2} \in C_i$ lying in general position is best represented by the codebook vector dictionary D_i belonging to C_i . This intuitive assumption translates into the following class label assignment for any data vector $Y \in R^{NP^2}$:

$$y \in C_i, C_i = \arg \min_i \|y - d_i x_i\|^2$$

where x_i is the sparse code for the data vector y .

A similar approach is used in automatic face recognition by [13]. But the proposed idea is different from [13] as they use l_1 minimization and use the training vectors directly as the dictionary atoms. In the algorithm presented below D_{ad} , D_{cn} are the dictionaries of AD and CN obtained, after applying the KSVD algorithm by using their data respectively. The pseudo code for classification between AD and CN classes is given the following listing:

Algorithm for Error Representation between CN and AD

1. Compute X_{cn} using $\text{OMP}(D_{cn}, y, 2)$
2. Compute X_{ad} using $\text{OMP}(D_{ad}, y, 2)$
3. Error representation of CN $\text{error}_{CN} = \|y - D_{cn} X_{cn}\|_2^2$
4. Error representation of AD $\text{error}_{AD} = \|y - D_{ad} X_{ad}\|_2^2$
5. if $\text{error}_{CN} < \text{error}_{AD}$
6. Then 'Y belongs to Normal'.
7. Else 'Y belongs to AD'.

4. RESULTS COMPARISION

By using the SVM the experimental results show that it achieves an accuracy of 90.32% for CN Vs AD, 89.65% for AD Vs MCI classification and an accuracy of 86.6% for MCI Vs CN classification, respectively. By using the Dictionary method for sparse image representation such as K-SVD classifying 93.54% for AD Vs CN, 93.10% for AD Vs MCI and 90% for CN Vs MCI. The figure below represents the performance of K-SVD and SVM classifying between the AD, Cognitively Normal and Mild Cognitive Impairment. Sensitivity and specificity of each test are defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \text{Specificity} = \frac{TN}{TN + FP}$$

Where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negatives respectively [14]. These probabilities reveal the ability to detect AD, CN or MCI patterns thus, the best algorithm is the one that achieves the best specificity and sensitivity. The results of both the methods were presented in the table 2 and table 3.

Table-2: Performance figures of SVM classifier

SVM	CN Vs MCI	CN Vs AD	MCI Vs AD
Accuracy	86.67%	90.32%	89.65%
Sensitivity	83.87%	87.50%	86.66%
Specificity	66.66%	75.00%	75.00%

Table-3: Performance figures of K-SVD classifier

K-SVD	CN Vs MCI	CN Vs AD	MCI Vs AD
Accuracy	90.00%	93.54%	93.10%
Sensitivity	87.67%	89.50%	89.50%
Specificity	80.00%	75.00%	100.00%

5. CONCLUSIONS

In this paper, we have investigated using the K-SVD for classification of high dimensional MRI data. Furthermore, we have presented a random patch based classification with the K-SVD. The local patches are extracted from the relevant regions to capture the local spatial consistency and are randomly sampled to construct a feature subspace. Then, multiple classifiers are combined to make more accurate and robust classification. The experimental results on ADNI database show that K-SVD achieves better classification performance than SVM classifier when features are used for classification. In the current paper, we validate our method using MRI data from ADNI. However, other modality of data can also be used in our method.

REFERENCES

- [1]. Herrera L.J., Rojas I., Pomares H., Guillen A., Valenzuela O. and Banos O., "Classification of MRI Images for Alzheimer's Disease Detection" 2013 IEEE Conference, 2013, Page(s): 846 - 851.
- [2]. Wells. W.M, Grimson, W.E.L. Kikinis. R, and Jolesz. F.A "Adaptive segmentation of MRI data" medical imaging, IEEE Transactions on vol 15, issue 4,1996, Page(s): 429- 442.
- [3]. Prashanthi Vemuri, Jennifer L. Whitwell, Kejal Kantarci and Keith A. Josephs, "Antemortem MRI based Structural Abnormality Index (STAND)-Scores Correlate with Postmortem Braak Neurofibrillary Tangle Stage" MST 2008 Neuroimage 2008 August 15; 42(2): 559-567.
- [4]. Remi Cuingnet, Emilie Gerardin and Olivier Colliot "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database" Neuroimage 2010, YNIMG-07400, pages-16
- [5]. Stefan Kloppel, Cynthia M. Stonnington, Carlton Chu, Bogdan Draganski, Rachael I. Scahill, Jonathan D. Rohrer, Nick C. Fox, Clifford R. Jack Jr, John Ashburner and Richard S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease" brain oxford journals, 2008, 131, 681-689
- [6]. Yifeng Li and Alioune Ngom "Fast Sparse Representation Approaches for the Classification of High-Dimensional Biological Data" School of Computer Science, University of Windsor, Windsor, Ontario, Canada. 2012, Page(s): 1- 6
- [7]. Manhua Liu, Daoqiang Zhang, Dinggang Shen and the Alzheimer's disease Neuroimaging Initiative "Ensemble Sparse Classification of Alzheimer's Disease" 2012 April 2; 60(2): 1106-1116
- [8]. Shenghua Gao, Ivor Wai-Hung Tsang, and Liang-Tien Chia "Sparse Representation With Kernels" IEEE transactions on Image processing, VOL. 22, NO. 2, 2013
- [9]. Philip Breen "Algorithms for Sparse Approximation" Project, School of Mathematics University of Edinburgh 2009
- [10]. Rubinstein R., Bruckstein A.M. and Elad M. "Dictionaries for sparse representation modeling" IEEE Publication Year: 2010, Page(s): 1045 - 1057.
- [11]. Hien Van Nguyen, Vishal M. Patel, Nasser M. Nasrabadi, and Rama Chellappa "Kernel dictionary learning" UMIACS, University of Maryland, College Park, MD U.S. Army Research Laboratory, Adelphi, 2012, Page(s): 2021-2024
- [12]. Michal Aharon, Michael Elad, and Alfred Bruckstein "K-SVD: An Algorithm for Designing Over complete Dictionaries for Sparse Representation" IEEE Transactions on Signal Processing, VOL. 54, NO. 11, NOV 2006
- [13]. John Wright, Allen Y. Yang and Arvind Ganesh "Robust Face Recognition via Sparse Representation" IEEE transactions on pattern analysis and machine intelligence, VOL. 31, NO. 2, FEBRUARY 2009
- [14]. F. Segovia, J. M. Gorriz, J. Ramirez, D. Salas Gonzalez, I. Alvarez, M. Lopez, and R. Chaves, "A comparative study of feature extraction methods for the diagnosis of Alzheimer's

disease using the ADNI database" The Alzheimer's Disease Neuro imaging Initiative Neuro computing 75(2012)64-71

BIOGRAPHIES



Research Scholar at NIT Calicut, Kerala, India. M.Tech from Amrita Vishwa Vidyapeetham, Coimbatore, India



Pursuing M.Tech from NIT Calicut, Kerala, India
B.Tech: PVPSIT, Vijayawada, A.P, India



Professor of Computer Engineering and Dean Academic NIT Calicut, India. PhD IISC Bangalore