Journal of Neuroscience Methods 199 (2011) 140–145

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

MRI-based age prediction using hidden Markov models

Bing Wang, Tuan D. Pham*

Bioinformatics Research Group, School of Engineering and Information Technology, The University of New South Wales, Canberra ACT 2600, Australia

ARTICLE INFO

Article history:
Received 10 January 2011
Received in revised form 12 April 2011
Accepted 14 April 2011

Keywords:
MRI
Age prediction
Hidden Markov models
Wavelet transforms
Vector quantization
Kullback–Leibler divergence

ABSTRACT

Cortical thinning and intracortical gray matter volume losses are widely observed in normal ageing, while the decreasing rate of the volume loss in subjects with neurodegenerative disorders such as Alzheimer's disease is reported to be faster than the average speed. Therefore, neurodegenerative disease is considered as accelerated ageing. Accurate detection of accelerated ageing based on the magnetic resonance imaging (MRI) of the brain is a relatively new direction of research in computational neuroscience as it has the potential to offer positive clinical outcome through early intervention. In order to capture the faster structural alterations in the brain with ageing, we propose in this paper a computational approach for modelling the MRI-based structure of the brain using the framework of hidden Markov models, which can be utilized for age prediction. Experiments were carried out on healthy subjects to validate its accuracy and its robustness. The results have shown its ability of predicting the brain age with an average normalized age-gap error of two to three years, which is superior to several recently developed methods for brain age prediction.

1. Introduction

Age-related changes in brain morphology, including cortical thinning and gray matter (GM) atrophy, have been widely observed in ageing people. These brain changes can be firstly observed on the magnetic resonance imaging (MRI) in middle age (Salat et al., 2004; Chan et al., 2003). These alternations in the brain follow certain patterns. The morphological changes may be accelerated in some specific brain regions. An example includes the prefrontal cortex where age related changes can be greater than in other regions (Sowell et al., 2003). Meanwhile, neurodegenerative diseases, such as Alzheimer's disease (AD) and dementia accelerate brain tissue loss at a faster rate than the normal brain ageing process. Furthermore, recent reports (Vivek et al., 2006; Spulber et al., 2010) have revealed that brain tissue atrophy caused by AD is more regionally specific than normal ageing. These two factors exhibit properties that are distinguishable from the normal ageing brain morphology. Therefore, it is possible to estimate the stages of ageing according to the morphology observed on MRI data. The ability to predict deviations in brain morphology, from the normal ageing pattern, before the pathological onset has the potential of improving clinical diagnosis and treatment by early intervention (Christos et al., 2009; Spulber et al., 2010; Sluimer et al., 2009; Driscoll et al., 2009; Fotenos et al., 2008).

Several studies have been directed at detecting accelerated ageing in the brain: if the predicted age according to brain images is older than the subject's real age, then this could be evidence of fast ageing. It has been realised that in order to identify faster brain atrophy, the construction of a healthy MRI-based brain ageing model is required to validate the accuracy and robustness of the prediction (Franke et al., 2010).

There are four exploratory modelling methods for modelling neuronal ageing using MRI data. The first is support vector machines (SVM) (Lao et al., 2004) which assigns subjects into four age stages. The second is relevance vector machine for regression (RVR) based age prediction using principal component analysis for feature selection (Franke et al., 2010). The method was applied to 550 healthy subjects and a mean absolute error of 4.98 years was obtained. In (Ashburner, 2007), age prediction was estimated by using RVR, which yielded a root mean squared error of 6.5 years. The fourth example is quantitative brain water maps (BWM) which predicts brain age with a median absolute deviation of 6.3 years between real and predicted ages (Neeb et al., 2006).

All of these prediction methods are based on high dimensional morphological analysis (HDMA) (Fan et al., 2007). Although the HDMA approach has been widely applied, it has several technical challenges, including the need for large training data and effective feature selection. In order to overcome the limitation of the HDMA, and to detect subtle structural changes of the brain for identifying accelerated ageing, we propose to build a structural brain model for each subject by using the framework of hidden Markov models (HMMs). We then estimate the brain age of a target subject by com-

* Corresponding author. Tel.: +61 2 62688410.
E-mail addresses: bing.wang@student.adfa.edu.au (B. Wang), t.pham@adfa.edu.au (T.D. Pham).

0165-0270/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2011.04.022

puting the similarity between the constructed HMM of the subjects by using the Kullback–Leibler divergence.

The rest of the paper is organised as follows. In Section 2, we briefly present the mathematical components for building an HMM of the MRI-based brain structural model. Section 3 presents the concept of the Kullback–Leibler divergence as a basis for the HMM-based similarity analysis. Section 4 illustrates and discusses the performance of the proposed approach using a public MRI database. Finally, Section 5 is the summary of the findings, and highlights of some suggested issues for further study

2. HMM for MRI-based brain structural model

In order to construct an HMM for the brain model, we extract a useful feature of the MRI of the brain using the wavelet coefficients. The wavelet coefficients are then coded using a vector quantization technique for modelling the hidden states of the HMM. We also discuss other components of the HMM and finally provide a summary of the computational procedure for building the HMM for the MRI-based brain structural model.

2.1. Wavelet-feature extraction and symbol coding

The wavelet transform is a widely used signal processing tool which is often applied in multisolution analysis and image compression. It is capable of capturing data features based on different frequency bands. An example includes a 2D image: vertical, horizontal spatial frequency characteristics of the image. This procedure intuitively decomposes images into a number of scales, each of which represents a certain coarseness of the data under study. The discrete wavelet transform (DWT) is commonly used for signal analysis that decomposes original signals into approximation and oscillation (detail) coefficients. Approximation is a basic and smooth representation of the original signal, while oscillation provides the high frequency components within the original signal.

In brain MRI, DWT can be used to extract basic voxel intensity distribution patterns to characterise the distribution patterns of the brain tissue and CSF, while discarding the high frequency oscillations. The high frequency components may contain noise that is introduced by the registration and normalization steps, which human eyes cannot distinguish. The DWT can be applied on different scales of the brain MRI; for example, in each block of scattered MRI slices, on each slice of MRI or the whole volume. We are interested in exploring the local structure of the brain tissues and CSF. We also aim to capture the gradual structural changes within the brain, accompanied with ageing. Therefore, we implemented the DWT in each block of scattered MRI slices of every subject.

In the present work, only GM, WM and CSF regions of the brain MRI data were used for the wavelet feature extraction, as our purpose is to detect the structural and volume changes of different tissues rather than the intensity or density alterations. We divided each axial slice of a brain MR image into $N \times N$ blocks, rearranged the voxel values of each block into a vector to extract its wavelet approximation. The approximation of the DWT reflects how different tissues are spatially distributed in each block of the MR images. For example, when the atrophy of GM exists in one block, the position of different tissue indices would alter. Thus, the brain structural changes with ageing can be captured by this feature, and further reflected in the brain model we are going to build. In order to avoid the effect of noise and registration errors in the detail coefficients of the wavelet transform, we only preserved the wavelet approximation coefficients.

To implement the wavelet transforms, two sets of functions are involved in the procedure: scaling and wavelet functions representing low and high pass filters, respectively. The Daubechies wavelet transform was applied in our study. The high-pass and low-pass functions are given as follows.

$$y_{\text{high}}[k] = \sum_n x[n]g[2k-n] \tag{1}$$

$$y_{\text{low}}[k] = \sum_n x[n]h[2k-n] \tag{2}$$

where y_{high}, y_{low} are the outputs of the highpass g and lowpass h filters after resampling by two. The procedure can be repeatedly applied to the approximation of the last scale to produce approximation and detail on a coarser scale until a desired level is reached which forms a pyramidal structure (Mallat, 1989).

The wavelet coefficient vectors defined above are expected to be able to extract substantial information of the strcutures of the tissues and CSF. However, since we extract these vectors from each block where there is inevitable spatial relevance leading to redundancy. A data reduction step is required and discussed in the following procedure using a vector quantization (VQ) technique. Vector quantization is a data compression method, which utilizes codevectors to represent the source vectors in their proximity. VQ can reduce the amount of data, storage requirement and computational complexity. A set of codevectors which best represent a training dataset is called a codebook. Suppose we have a group of M source vectors. $\mathbf{T} = \{x_1, x_2, \ldots, x_M\}$, every vector is k-dimensional. $\mathbf{x_m} = (x_{m1}, x_{m2}, \ldots, x_{mk})$, $m = 1, 2, \ldots, M$. Let N be the number of codevectors and $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c_N}\}$, every $\mathbf{c_n}$ is k-dimensional as $\mathbf{x_m} : \mathbf{c_n} = (c_{n1}, c_{n2}, \ldots, c_{nk})$, $n = 1, 2, \ldots, N$. Let S_n represent the encoding space associated with code vector $\mathbf{C_n}$, $P = \{S_1, S_2, \ldots, S_N\}$ is the partition of the encoding space. If a source vector x_m is located in an encoding space S_n, then its approximation is C_n, denoted by $Q(\mathbf{x_m}) = c_n$. The average distortion is given by:

$$D_{\text{VQ}} = \frac{1}{MK} \sum_{t=1}^{M} (||\mathbf{x}_t - Q(\mathbf{x}_t)||)^2 \tag{3}$$

The VQ process can be shortly described as follows: given $\mathbf{T}$ and N, find $\mathbf{C}$ and P, such that D_{VQ} is minimized. However, $\mathbf{C}$ and P must follow the two criteria (Gersho and Gray, 1992). One is the nearest neighbour condition: the encoding region S_n consists of all vectors that are closer to $\mathbf{c_n}$ than any of the other codevectors. For example, if one vector is determined to belong to region S_n, then the distance between the vector and the center of S_n should be shorter than any distances between the vector and the centres of other regions. For vectors that are on the boundary of any region, a tie-breaking procedure can be applied to determine which region these vectors belong to. The other criterion is the centroid condition: codevector $\mathbf{c_n}$ should be the average of all training vectors that are in encoding region S_n. The most commonly used VQ method is the LBG algorithm (Gray, 1984). The LGB starts with an initial codebook, and then iteratively splits the training data into two codevectors until the desired number of the codevectors is reached. The data to be quantized in this study are the wavelet coefficient vectors which are extracted from the MRI data of each subject.

In order to compare the similarity between subjects, the code vectors should be consistent from subject to subject. However, the more the number of subjects increases, the more the computer memory requires. In order to avoid the problem of large computational load in the VQ design, we propose to build the same codebook for every two subjects for all the pairwise combinations. In other words, the codevectors of every two subjects are pooled together for the construction of the codebook, and then based on the feature (wavelet) vectors of each subject, the state transitions of each subject can be obtained from the pooled codebook. The number of code vectors was experimentally chosen to be 32, which is among