Computer Methods and Programs in Biomedicine 151 (2017) 203–212

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Brain parcellation based on information theory

Ester Bonmati, Anton Bardera, Imma Boada

Institute of Informatics and Applications, University of Girona, Campus Montilivi, 17003 Girona, Spain

ARTICLE INFO

Article history:
Received 10 November 2015
Revised 10 April 2017
Accepted 31 July 2017

Keywords:
Brain parcellation
Hierarchical clustering
Mutual information
Human brain connectome
Markov process

ABSTRACT

Background and objective: In computational neuroimaging, brain parcellation methods subdivide the brain into individual regions that can be used to build a network to study its structure and function. Using anatomical or functional connectivity, hierarchical clustering methods aim to offer a meaningful parcellation of the brain at each level of granularity. However, some of these methods have been only applied to small regions and strongly depend on the similarity measure used to merge regions. The aim of this work is to present a robust whole-brain hierarchical parcellation that preserves the global structure of the network.

Methods: Brain regions are modeled as a random walk on the connectome. From this model, a Markov process is derived, where the different nodes represent brain regions and in which the structure can be quantified. Functional or anatomical brain regions are clustered by using an agglomerative information bottleneck method that minimizes the overall loss of information of the structure by using mutual information as a similarity measure.

Results: The method is tested with synthetic models, structural and functional human connectomes and is compared with the classic *k*-means. Results show that the parcellated networks preserve the main properties and are consistent across subjects.

Conclusion: This work provides a new framework to study the human connectome using functional or anatomical connectivity at different levels.

1. Introduction

The human brain contains an extraordinary network of roughly one hundred billion neurons capable of sharing and processing information efficiently. The *connectome* models these connections as a graph, where nodes represent brain areas and edges represent structural or functional connections [1,2]. To define the nodes, parcellation methods are used to subdivide the brain cortex into different regions according to a predefined criterion (i.e., cytoarchitecture, structure, function...).

Atlas-based parcellation methods subdivide the brain by employing a three-dimensional anatomical template [3]. This template can be based on cytoarchitecture, electrophysiological observations, cortical curvature patterns [4], structural or functional connectivity profiles [5], among others. A limitation of these methods is the lack of individuality as they are based on a sample dataset as opposed to the subject.

On the other hand, *connectivity-based* parcellation methods subdivide the brain into spatially coherent regions of homogeneous connectivity by grouping grey-matter voxels according to the similarity of their connectivity patterns [3], obtained from diffusion magnetic resonance imaging (dMRI) or functional magnetic resonance imaging (fMRI). The more popular methods are based on the *k*-means approach [6–8], which groups voxels into *k* non-overlapping clusters using a similarity measure. The main issues with these methods are the definition of the number of clusters a priori and the reliance on initial random sampling, as it has been shown that iterative repetitions of the same method may lead to different results [9]. To overcome these limitations, and assuming that brain networks have hierarchical properties [10,11], several hierarchical clustering methods that compute a parcellation at each level in the hierarchy have been proposed [12–15]. These methods obtain brain parcellations at multiple granularities without the need to define the number of clusters.

Connectivity-based methods are strongly dependent on the similarity measure used by the algorithm. Gorbach et al. [13] proposed a hierarchical method that clusters voxels using the mutual information between tractograms, and obtained promising results for specific regions of the brain. The use of mutual information as a similarity measure is, therefore, an effective solution to group voxels. However, Gorbach et al.'s method [13] assumes that the whole cluster can be represented by only one tractogram.

E-mail address: ester.bonmati@imae.udg.edu (E. Bonmati).

http://dx.doi.org/10.1016/j.cmpb.2017.07.012
0169-2607/

1.1. Our approach

In this paper, we present a hierarchical parcellation method that preserves the structure of brain network with no need to define representative tractograms. We model brain networks as a random walk on the connectome by using the structural or functional connectivity matrix. From this model, we quantify the brain structure. Brain regions are clustered by applying a bottom-up hierarchical method based on the information bottleneck using a control process [16]. We evaluate the parcellation method by using synthetic, structural, and functional brain networks at different scales. The robustness of the method is tested by doing multiple-subject comparisons with the resulting vector of hierarchical clusters.

2. Method

In this section, we propose a new method to parcellate the brain. First, we introduce a brain model based on a Markov process. Then, we describe the parcellation method, which uses the information bottleneck-based method. Additionally, we describe a measure based on mutual information which is used to perform pairwise group comparisons.

2.1. Markov process-based brain model

A brain network can be modeled as a graph with a pair of sets $G = (V, E)$, where V represents the set of v brain regions, denoted by $\{V_1, \ldots, V_v\}$, and E the set of e edges between two nodes of V, that denotes their anatomical or functional connectivity. This graph can be represented by a connectivity matrix C with $v \times v$ elements, where C_{ij} gives the connectivity weight between node i and node j.

In this paper, we propose to model a brain network as a Markov process $\mathbf{X} = \{X_0, X_1, \ldots, X_t, X_{t+1}, \ldots\}$, which represents a random walk of a particle moving from one brain region to another. From this model, we can define a probability density function $p(X_t) = \{p(x_1^t), \ldots, p(x_i^t), \ldots, p(x_v^t)\}$, where $p(x_i^t)$ represents the probability that a particle takes the value x_i (i.e. the particle is in the brain region i) at state X_t (i.e. at time step t). This particle randomly moves from a node x_i to node x_j according to the connectivity or probability defined in the transition probability matrix, whose elements are given by

$$p(x_j^{t+1}|x_i^t) = \frac{C_{ij}}{C_i}, \tag{1}$$

where $C_{ij} = C_{ji}, \forall i, j$ and $C_i = \sum_i C_{ij}$ is the total weight of the edges emanating from node x_i. The transition probability $p(x_j^{t+1}|x_i^t)$ defines the probability of being in node x_j after visiting node x_i. Note that the transition probability depends only on the current state and not on the previous ones.

The transition probabilities can be used to define the transition distribution from each node x_i, which is given by

$$\begin{aligned} p(X_{t+1}|x_i^t) &= \{p(x_1^{t+1}|x_i^t), \ldots, p(x_j^{t+1}|x_i^t), \ldots, p(x_v^{t+1}|x_i^t)\} \\ &= \left\{\frac{C_{i1}}{C_i}, \ldots, \frac{C_{ij}}{C_i}, \ldots, \frac{C_{iv}}{C_i}\right\}. \end{aligned} \tag{2}$$

This distribution represents the overall probability of a particle to be in a different node after visiting node x_i.

The probability of being in node x_i can be defined by a stationary distribution [17]. In this case, for undirected brain networks, the stationary distribution is given by

$$p(x_i) = \frac{C_i}{C_T}, \tag{3}$$

where $C_T = \sum_i \sum_j C_{ij}$ is twice the sum of the weights of all the edges [17]. The stationary distribution, $p(X_t) = \{p(x_1), \ldots, p(x_i), \ldots, p(x_v)\}$, defines the probability of a particle to be in each of the nodes. Note that the stationary distribution of a node is proportional to the total weight of the edges emanating from that node.

2.2. Mutual information as a measure of brain structure

Mutual information (MI) is a well-known measure that quantifies the shared information between two different variables X and Y defined as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \end{aligned} \tag{4}$$

where $H(X) = -\sum_{x\in\mathcal{X}} p(x) \log p(x)$ is the Shannon entropy of X and measures the uncertainty of the variable X, and $H(X|Y) = -\sum_{y\in\mathcal{Y}} p(y) \sum_{x\in\mathcal{X}} p(x|y) \log p(x|y)$ is the conditional entropy and measures the average uncertainty associated with X if we know the outcome of Y. In our approach, we use the MI measure to quantify the shared information or similarity between two states of a Markov chain, i.e. $I(X_t; X_{t+1})$.

For a stationary Markov chain, the MI between consecutive states, $I(X_t; X_{t+1})$, coincides with the *excess entropy* [18,19], which is a measure of system structure. We use this measure to quantify the structure of the networks. High values of MI will indicate that there is a high correlation between consecutive states and, therefore, that the brain is highly structured. Mutual information can also be seen as the difference between the uncertainty of the states without any prior knowledge and the uncertainty of the states when the past is known (or information gained when the previous node is known). Therefore, the higher the MI, the less random the connections.

2.3. Parcellation method

The goal of our parcellation method is to cluster the brain regions, represented as different states of a Markov chain, by minimizing the loss of information when two regions are merged, the effect of which is to maintain the overall structure. The *agglomerative information bottleneck method*, proposed by Thisby et al. [20], clusters a random variable X depending on a random variable Y by minimizing the loss of mutual information. Gorbach et al. [13] used this method to preserve the maximum information between the representative tractogram and all the tractograms that belonged to the same cluster. However, in our brain model, we want to preserve the maximum information between brain regions, represented as consecutive states (X_t and X_{t+1}) of a Markov process $\mathbf{X}$ instead of two different variables X and Y. In this case, when two nodes of the Markov process are merged, both X_t and X_{t+1} are modified. Due to this limitation, the classic agglomerative information bottleneck method cannot be used. Thus, we use the extended version of the algorithm presented in [16], which takes this fact into consideration and enables to cluster a random variable X depending on a Markov process.

The method starts by assigning each node (or brain region) to a different cluster, with a total of v clusters. In the first step, the loss of mutual information due to a possible merge of every pair of nodes (x_i, x_j) is calculated. The MI loss is computed as the difference of mutual information between two consecutive states $I(X_t; X_{t+1})$, when nodes x_i and x_j belong to different clusters, and the mutual information $I(\widehat{X}_t; \widehat{X}_{t+1})$ when nodes x_i and x_j have been