

Functional Neuroimaging Data Mining Radu Mutihac*

Department of Physics, University of Bucharest, Bucharest 077125, Romania

*Corresponding author: Radu Mutihac

✉ mutihac@gmail.com

Department of Physics, University of Bucharest, Bucharest 077125, Romania.

Tel: +4072-702-0772

Fax: +4021-315-9249

Abstract

Data mining, alternatively denominated as knowledge discovery from data, is a relatively young and fast-growing interdisciplinary scientific field. The contribution hereafter critically underpins the main approaches and trends in data mining applied to functional neuroimaging analysis.

Keywords: Data mining; Exploratory data analysis; Cluster analysis; Multivariate regression; Functional brain imaging; Statistical parametric mapping

Received: March 21, 2018; **Accepted:** May 07, 2018; **Published:** May 14, 2018

Introduction

Exploring physical world is aiming to discover the structure in experimental data and to reveal the underlying processes from which the acquired data have originated. In most practical cases, useful information comes out by processing raw data only. Besides, real-life measurements provide unknown mixtures of interesting and uninteresting signals. A popular saying goes like we are living in the information age", yet we are actually living in the data age" [1]. And still, data mining might be regarded as a direct development of information technology looking for novel approaches in data processing.

Functional neuroimaging data yield valuable information on physiological processes, yet being often affected by several artifacts including noise, and generally acquired as collections of unknown mixtures of signals variably summing up in time and/or space. In several cases, even the nature of signal sources is a question of debate. In most cases, reliable and full information is missing, so that a reasonable estimation of plausible solutions to identification of original signals falls in the large class of blind source separation (BSS) methods [2].

Fundamental Concepts

Some fundamental concepts in data analysis are loosely defined hereafter in order to introduce the general context of our discussion.

Computer Science (CS) though controversial debates exist on the meaning of "computer science", we accept CS as the scientific discipline and practical approach to computation and its applications [3]. It entails a systematic exploration of the feasibility, structure, expression, and formal reasoning of the protocols that reflect the acquisition, representation, processing, storage, communication of, and access to information, either

if information is encoded as bits in a computer memory or transcribed engines and protein structures in a human cell. Nevertheless, disputes have etymologically stemmed, in the sense that there is "no science" in CS as far as CS is not concerned with observing nature. In this respect, parts of CS are engineering (more practical), and parts are mathematics (more theoretical). Contrarily, science refers to laws of nature and natural phenomena, whereas phenomena involved in CS are man-made [4]. Apart from a consistent definition unanimously agreed upon, CS is commonly accepted to consist of some theoretical and practical subdomains, as follows.

- Computational complexity theory: addressing fundamentals of computational and intractable problems; it is highly abstract;
- Computer graphics: dealing with real-world computer-assisted visual applications;
- Programming language theory: investigating analytical approaches to programming;
- Computer programming: exploring programming languages and complex systems;
- Human-computer interaction: designing usable computers and computations universally accessible to humans.

The field of *Artificial Intelligence* (AI) has emerged and developed assuming that a specific property of humans, that is, intelligence, can be sufficiently well de- scribed to the extent that it can be mimicked by a machine. As such, philosophical issues arise on the nature of the human mind and, furthermore, on the

ethics of creating artificial systems endowed with human-like intelligence, issues which have been addressed by myth, fiction, and philosophy since antiquity [5]. AI has been the subject of optimism by its inception, still it has also crossed setbacks ever since. For the time being, AI constitutes a major component of technology and poses several challenging problems at the forefront of research in CS.

Mechanical or formal reasoning has been introduced by philosophers and mathematicians as well since antiquity, too. The study of logic has established a milestone in our society by the creation of the digital electronic computer. The starting point was marked by the Turing machine [6]. Turing's theory of computation demonstrated that a programmable machine may simulate any act of mathematical deduction by manipulating simple symbols like "0" and "1". By the same time, discoveries in cybernetics, information theory, and significant advances in neurology, oriented the interest of the scientific community towards evaluating the feasibility of designing an electronic brain.

Data Mining (DM) equates to the extraction of implicit, not *a priori* known, and potentially valuable information from raw data [7]. The underlying idea in DM is to build up computer programs that seeks for regularities or patterns through databases. Anyway, real data are imperfect, incomplete, corrupted, contingent on accidental coincidences, and some of no interest whatsoever, leading to spurious and inexact predictions. Some exceptions will still exist to all rules, as well as cases not covered by any rule. Therefore, the algorithms involved in DM must be robust enough to cope with imperfect data, yet capable to identify inexact but useful regularities [8]. As an analytic process, DM is conceptually designed to work in three stages as it follows hereafter.

- a) Initial exploration of data in search of consistent patterns, as well as systematic relationships among variables. The process may involve data cleaning, data transformations, and data space reduction to a proper subspace by feature selection, thereby reducing the number of variables to a meaningful and manageable range,
- b) Developing models based on pattern identification and statistical assessment of findings by means of detected patterns applied to new data subsets. Ranking models and choosing the best one based on their predictive performance, that is, explaining the variability in data and providing consistent results across samples,
- c) Deployment and prediction by using the model previously chosen as best to disclose knowledge structures aiming to guide further decisions under conditions of limited (if any) accessibility and/or certainty.

Machine Learning (ML) is conceptualized as the technical basis of DM aiming to discover and describe structural patterns in data. In other words, ML is perceived as the acquisition of structural descriptions from examples, further employed for prediction, explanation, and understanding. Historically, ML was considered by Arthur Samuel the subfield of Computer Science (CS) that gives computers the ability to learn without being explicitly programmed" [9]. Specifically, ML refers to the construction

and study of computer algorithms that are automatically self-improving through experience in that they can learn and make predictions on data. Such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. The core of ML deals with representation and generalization. Feature learning or representation learning is a set of techniques that learn a transformation of "raw" inputs to a representation that can be effectively further exploited in supervised/unsupervised learning tasks. Generalization is the property that the system will perform well on unseen data instances; the conditions under which this can be guaranteed are a key object of study in the subfield of computational learning theory. Contrarily, traditional statistical techniques are not adaptive but typically process all training data simultaneously before being used with new data. In a similar key, Tom Mitchell, another learning researcher, proposed back in 1977 a more precise definition for ML in the case of well-posed learning problem: "A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E " [10]. In a broader view, ML is closely related to and overlaps with *computational statistics*, which focuses on prediction-making through the use of computers, too. It is also related with *mathematical optimization*, which provides theory, methods, and application domains to the field. ML is sometimes conflated with DM regarding prediction, although the latter overlaps more on exploratory data analysis (EDA) sharing much in common in terms of goal and methods.

Methods

In practice, raw data *per se* are generally of little (if any) immediate usage. It is only when information is extracted via processing that makes data meaningful.

Data analysis

Back in '60s, Tukey advocated that classical statistics leaning on analyzing small, homogeneous, stationary data by means of known distributional models and assumptions will prove inappropriate to deal with the problems raised by the analysis of large amount and complex data" [11]. The reason invoked was the qualitative difference that might exist between practical larger and larger data sets at hand and common smaller ones rather than strictly the size [12]. Consequently, functional neuroimaging data analysis should primarily rely on methods circumscribed to both DM and EDA. Moreover, Huber stated that ". . . there are no panaceas in data analysis" [12]. It comes out that an optimal choice is domain-dependent (**Figure 1**). In order to make it clear, mathematical problems are considered ill-posed if they do not satisfy each of the three criteria:

- a) a solution exists,
- b) it is unique, and
- c) it depends continuously of the initial data.

To solve ill-posed problems, well posed ness must be restored by restricting the class of admissible solutions [13].

Data mining and exploratory data analysis

DM is an iterative process of exploring and modeling large amounts of data aiming to discover baseline patterns and relationships among significant variables. As such, DM is called to identify trends, predict future events, and assess various courses of action that improve system performance. DM is a multidisciplinary field importing and boosting ideas and concepts from diverse scientific areas like statistics, signal and image processing, pattern recognition, mathematical optimization, and computer vision. Extracting non-explicit knowledge is fostered by advances in several disparate and often incongruous domains, such as bioinformatics, DNA sequencing, e-commerce, knowledge management remote sensing images, stock investment and prediction analysis, and real-time decision making [14].

The main applications of DM span a large range of issues in signal processing like adaptive system modeling and information mining including applications on biomedical data [15], visual data mining [16], scale-space analysis with applications in image segmentation [17], chemometrics including artificial neural networks (ANNs) [18], characterization of protein secondary structure [19], and many more.

Exploratory data analysis (EDA) consists of large set of techniques that deal with data informally and disclose structure quite straightforward. Data probing is primarily stressed upon, in many cases prior to their comparison with any particular probabilistic models. Such methods are optimal compromises in many circumstances and quite near to optimal solution for each individual case. In a typical exploratory approach, several variables are critically considered and thoroughly compared by means of diverse techniques in search of systematic patterns in data. In a more general sense, computational exploratory data analysis comprises various methods from a large spectrum ranging from simple basic statistics to advanced multivariate exploratory techniques. Basic statistical exploratory analysis includes techniques like:

- a) inspecting the distribution of variables,
- b) comparing the coefficients of the correlation matrices with meaningful thresholds, and
- c) inspecting multi-way frequency tables.

Some frequent approaches in multivariate exploration are listed hereafter.

- a) *Principal Component Analysis* (PCA) [20]
- b) *Independent Component (Subset) Analysis* (ICA) [21]
- c) *(Fuzzy) Cluster Analysis* (FCA) [22,11]
- d) *Factor Analysis* (FA) [23]
- e) *Projection Pursuit* (PP) [24]
- f) *Discriminant Function Analysis* (DFA)
- g) *Partial Least Squares* (PLS)
- h) *Multidimensional Scaling*
- i) *Log-linear Analysis*
- j) *Canonical Variate Analysis*
- k) *Correspondence Analysis*
- l) *Time Series Analysis*
- m) *Classification Trees*
- n) *Stepwise Linear and Nonlinear Regression*
- o) *Continuum Regression*
- p) *Multivariate Linear Model* (MLM)
- q) *General Linear Model* (GLM)

Analytical techniques encompass graphical data visualization techniques that can identify relations, trends, and biases usually hidden in unstructured data.

In functional brain imaging, EDA methods can identify interesting features reporting on brain activations which may be not anticipated or even missed by the investigator. Whereas EDA performs flexible searching for evidence in data, confirmatory data analysis (CDA) is concerned with evaluating the available evidence. In imaging Neuroscience, there is a permanent dynamic interplay between hypothesis generation on one hand, and hypothesis testing on the other hand, that can be regarded as a Hegelian synthesis of EDA and CDA [25]. Furthermore, artifactual behavior identified easily by EDA may raise questions on

1. data appropriateness,
2. the necessity of additional preprocessing steps, or
3. introduction of spurious effects by the preprocessing employed.

By all means, confirmatory methods are a must for controlling both type I (false positives) and type II (false negatives) errors, yet their statistical significance is meaningful if both the chosen model and the distributional assumptions are correct only.

DM is heavily based on statistical concepts including EDA

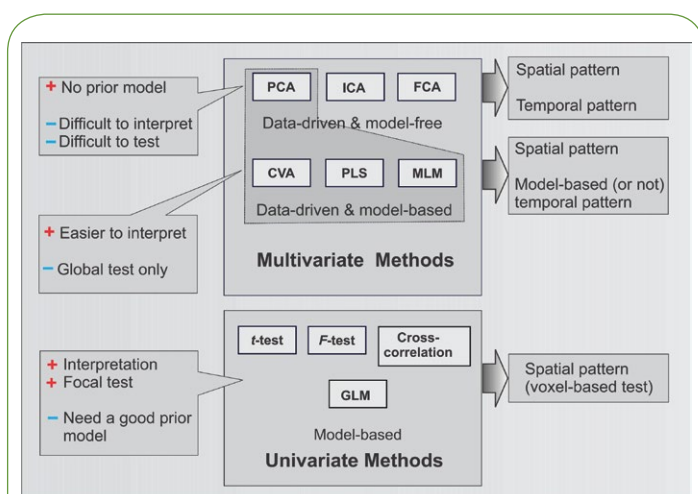


Figure 1 Hypothesis-driven versus model-driven univariate/multivariate data analysis with benefits (+) and pitfalls (-): PCA, FA, PP, CR, ICA, CA, FCA, CVA, PLS, MLM, GLM (abbreviations explained in text).

and modeling and, consequently, it shares with them some components. Nevertheless, an important difference exists in the goal and purpose between DM and traditional EDA in that DM is oriented towards applications to a larger extent rather than the underlying phenomena. That is, DM is less concerned with identifying the specific relations between the involved variables, rather its focus is on producing a solution that can generate useful predictions. Therefore, DM comprises traditional EDA techniques, as well as techniques like ANNs that can come out with valid predictions and still not having resources to identify the specific nature of the variable interrelations on which the predictions are made.

Exploratory methods

The central interest in functional brain studies resides in the electrical activity of neurons, which cannot be directly investigated by any magnetic resonance imaging (MRI) [26]. The human brain electrical activity is of paramount interest for both understanding and modeling the human brain, and for medical diagnosis and treatment as well, especially for developing automated patient monitoring, computer-aided diagnosis, and personalized therapy [26].

In data analysis, a widely spread task consists in finding an appropriate representation of multivariate data aiming to facilitate subsequent processing and interpretation. Transformed variables are hoped to be the underlying components, which best describe the intrinsic data structure and highlight the physical causes responsible for data generation. Linear transforms like PCA and ICA and are often envisaged to accomplish this task due to their computational and conceptual simplicity [26]. Generally, methods of unsupervised learning fall in the class of data-based (hypothesis-driven) analysis, such as eigen image analysis [27,28] or self-organizing maps (SOM) [29].

PCA and ICA: PCA is defined by the eigenvectors of the covariance matrix of the input data. In PCA, data are represented in an orthonormal basis determined by the second order statistics (covariances) of the input data. Such representation is adequate for Gaussian data [30]. PCA is a means of encoding second-order dependencies in data by rotating the orthogonal axes to correspond to the directions of maximum covariance (Figure 2). As a linear transform, PCA is optimal in terms of least mean square errors over all projections of a given dimensionality. PCA decorrelates the input data but does not address the high-order dependencies. Decorrelation means that variables cannot be predicted from each other using a linear predictor, yet nonlinear dependencies between them can still exist. Edges, as for instance, defined by phase alignment at multiple spatial scales, constitute an example of high-order dependency in an image, similarly to shape and curvature are [26]. Second-order statistics capture the amplitude spectrum of images but not the phase [31]. Coding mechanisms that are sensitive to phase play an important role in organizing a perceptual system [32]. The linear stationary PCA and ICA processes can be introduced on the basis of a common data model. The ICA model is a data-driven multivariate exploratory approach based on the covariance paradigm and formulated as a generative linear latent variables model [21]. ICA

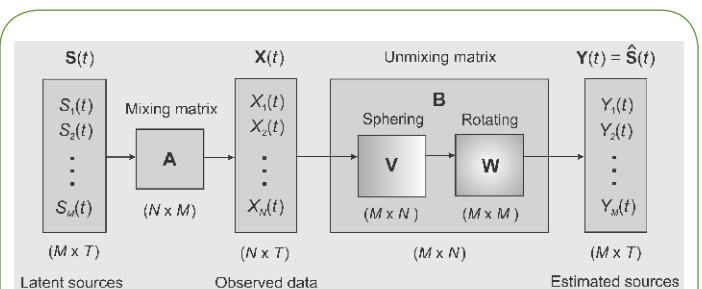
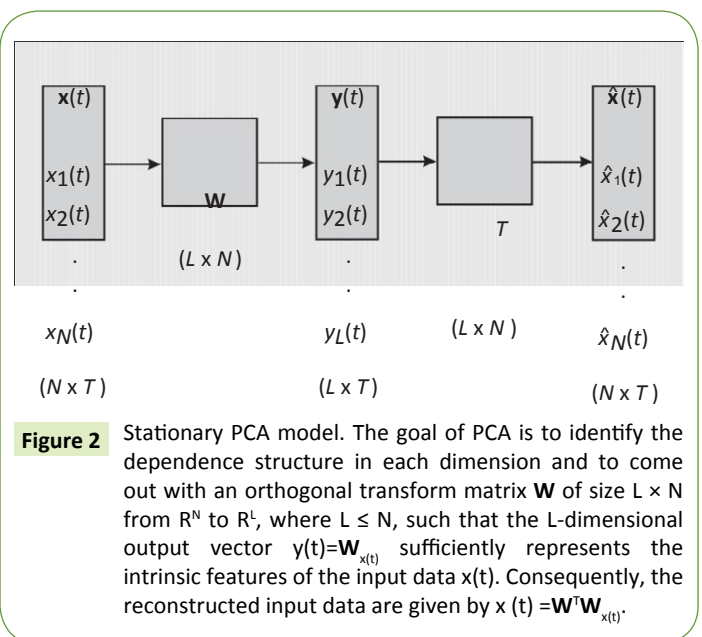
comes out with typical components like task-related, transiently task-related, and function-related activity without reference to any experimental protocol, including unanticipated or missed activations (Figure 3).

The assumptions behind the ICA model are the following:

- (i) the latent source signals are assumed statistically independent and at most one Gaussian, and
- (ii) the mixing process is assumed stationary and linear but unknown.

ICA, based on higher order-statistics, transforms the ill-posed PCA problem into a well-posed one. The ICA decomposition is unique up to IC amplitude (scale), IC polarity (sign), and IC ranking (order) [26]. Technically, applying the ICA model amounts to the selection of an estimation principle (objective function) plus an optimization algorithm. Typical objective functions consist in maximization or minimization of

- (i) high-order statistical moments (e.g., kurtosis),



- (ii) maximum likelihood (ML),
- (iii) mutual information (MI), or
- (iv) negentropy.

For ICA model, statistical properties (e.g., consistency, asymptotic variance, robustness) depend on the estimation principle. Algorithmic properties (e.g., convergence speed, memory requirements, numerical stability) depend on the selection of the optimization algorithm. ICA is based on the concept of *independence* between probability distributions which, in its turn, relies on information theory. *Entropy* is a criterion for statistical independence widely employed. Terms like information and entropy are richly evocative with multiple meanings in everyday usage; information theory captures only some of the many facets of the notion of information (Figure 4).

Clustering analysis: Searching for meaningful patterns in data like biological ones, has been a permanent endeavor best typified by the taxonomy that arranged species into groups based on their similarities and differences [33]. Clustering is an important DM method for discovering knowledge in multidimensional data. Clustering analysis, alternatively called automatic classification, numerical taxonomy, typological analysis, amounts to grouping, segmenting or partitioning a set of objects into subsets (clusters), maximizing their degree of similarity within each cluster and minimizing their degree of dissimilarity if belonging to distinct clusters. Clusters may be

- (i) disjoint vs. over-lapping,
- (ii) deterministic vs. probabilistic, and
- (iii) flat vs. hierarchical.

As for instance, *k*-means clusters are disjoint, deterministic, and flat.

There are four major approaches of clustering.

1. Hierarchical clustering: successive clusters are determined by means of previously established clusters;
2. Partitional (*k*-means) clustering: all clusters are iteratively determined simultaneously;
3. Model-based clustering;
4. Density-based clustering.

Hierarchical clustering is performed by

- (i) agglomerative methods (bottom-up), which merge the objects into successively larger clusters;
- (ii) divisive methods (top-down), which separate the objects into successively smaller clusters (Figure 5).

In general, the merges and splits are determined in a greedy manner. A greedy algorithm is an algorithmic paradigm that follows the problem-solving heuristic of making the locally optimal choice at each stage [34] and further looking for a global optimum. Greedy algorithms find the globally, optimal solution for some optimization problems, but may fail in several other instances. Clustering algorithms are attractive for class

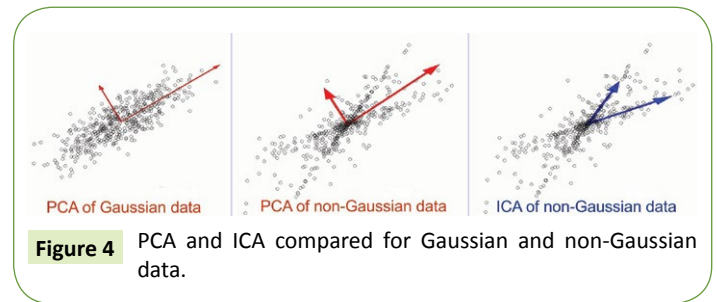


Figure 4 PCA and ICA compared for Gaussian and non-Gaussian data.

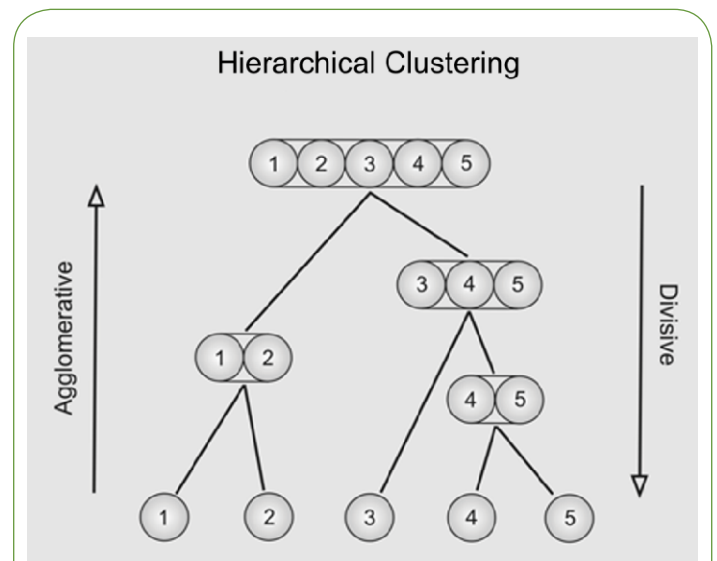


Figure 5 Hierarchical clustering. Agglomerative methods (bottom-up) merge the objects (observations) into successively larger clusters up to a single cluster grouping all objects. Divisive methods (top-down) separate the objects (observations) into successively smaller clusters down to clusters containing an object only.

identification in spatial databases but suffer from serious pitfalls when applied to large spatial databases [35].

The *hierarchical k*-means clustering is a hybrid approach for improving the performance of *k*-means algorithms. The *k*-means clustering method assumes that each data point is assigned to one cluster only [36]. Yet several practical situations suggest a *soft clustering* approach, where multiple cluster labels can be associated with a single data item, and each data point is assigned a probability of its association with more than one cluster [36]. A Gaussian mixture model with expectation-maximization algorithm (GMM-EM) typifies fuzzy cluster analysis (FCA) [4]. The fuzzy *c*-means (FCM) clustering generate fuzzy partitions and prototypes for any set of numerical data. The clustering criterion employed to aggregate subsets of data items consists in a generalized least-squares (GLS) objective function [37]. Most clustering algorithms may include some choices of distance measures like Euclidean, Manhattan, or Mahalanobis, an adjustable weighting factor that controls sensitivity to noise, acceptance of variable numbers of clusters, and outputs that include several measures of cluster validity [37]. Clustering by means of a symmetric or asymmetric distance measures makes an important distinction, which determines how the similarity of

two objects are calculated and, subsequently, their allocation to the same or different cluster. The results of hierarchical clustering are usually presented in dendrograms (Figure 6). Dendrograms are plotted after the calculation of a certain selected distance matrix and a linkage procedure has been opted for creating a picture on how merging can be employed to partition samples into subsets according to distance thresholds.

Partitional clustering decomposes directly a set of data items into a collection of disjoint clusters. Any criterion function to be minimized should stress on both local and global structure of data. In *K*-means clustering, the criterion function E_k is the average squared distance of the data items from their nearest cluster centroids. *K* centroids are initialized and subsequently their positions are adjusted iteratively by first assigning the objects to the nearest clusters and then recomputing the centroids.

The stopping criterion of iteration is enforced by E_k reaching insignificant further changes. Alternatively, each randomly chosen item may be successively considered and the nearest centroid consequently updated (Figure 7). The *K*-means algorithm is a most common approach and the simplest unsupervised learning algorithm to implement partitional clustering.

Model-based clustering assumes data as drawn from a distribution, which is a mixture of two or more clusters. Such approach finds best fitting models to available data and estimates the number of clusters [38].

Density-based clustering (DBSCAN) is a robust partitioning method introduced by Ester et al. back in 1996 [11]. It can discover clusters of arbitrary shapes and sizes from noisy data contaminated with outliers [38].

CA is a fundamental approach in unsupervised ML, including cancer research for classifying patients into groups according to their gene expression profile. As such, identifying the molecular profile of patients with good or bad prognostic, as well as elucidating the mechanisms of the disease itself can be carried out.

Whatever method or combination of methods would be used in exploratory analysis to discover new hypotheses (models) extracted directly from data, they have to be subsequently tested and verified by some more conventional statistical inferential methods of analysis. The combined information gathered from two or more methods may reveal structure in data that any single method could not have provided. The more an approach embeds prior knowledge that we are aware of about the structure to be discovered, the higher the chance of its detection. This suggests starting the analysis in an adequate Bayesian framework that incorporates all available information on the data and continuously updates the state of knowledge when new data are presented.

Functional neuroimaging

Functional MRI (fMRI), is a non-invasive neuroimaging modality, has emerged as a most convenient approach for mapping brain activated regions, both in health and in disease [39]. There are two most useful features of fMRI data sets fMRI signals are the characteristics, namely, *no stationarity* and *distributional heterogeneity* [40]. The analysis of large, quite complex, and heterogeneous functional neuro imaging data ought to start with an exploratory approach aiming to reveal the intrinsic

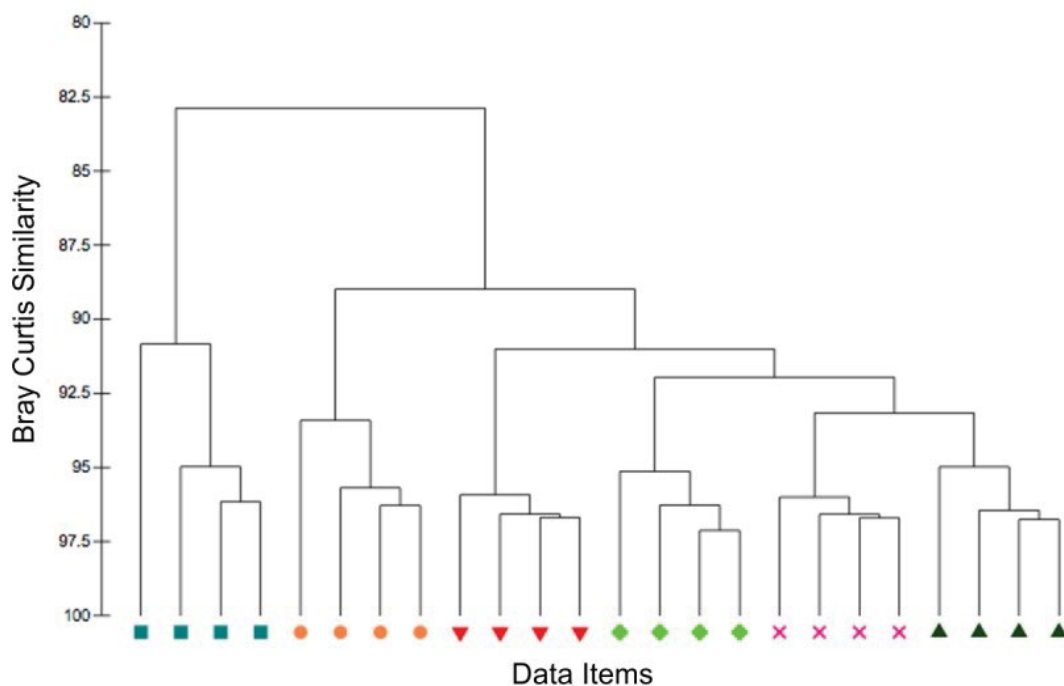


Figure 6 Dendrogram of cluster analysis based on Bray-Curtis similarity. Dendrogram shape is twofold arbitrary: the order on the x-axis is irrelevant as clusters can be rotated around any clustering distance leading to 2^{n-1} different sequences, and the distance matrix depends on the settings used.

structure in data with no need for prior models and minimal statistical assumptions, yielding as non-committal as possible results. Exploratory data-driven methods are complementary to hypothesis-led methods of confirmatory data analysis (CDA); the representative time courses that are the outcome may be conceptualized as alternative hypotheses to the null hypothesis H_0 (i.e., no activation) [41].

The hypothesis-driven (model-based) inferential methods used test specific hypotheses on the expected changes in blood oxygenation level dependence (BOLD) response. Such changes are specified as regressors in a multiple linear regression framework of the generalized linear model (GLM), and their relative weights are given by the regression coefficients (model parameters) like in **Figure 8** [41].

In any multisubject/multisession experiments, for a voxel to represent the same anatomical location for all subjects/sessions under all conditions, raw data are spatially/temporally and ultimately mapped into a standardized coordinate space that accounts for differences in brain size and orientation, such as the

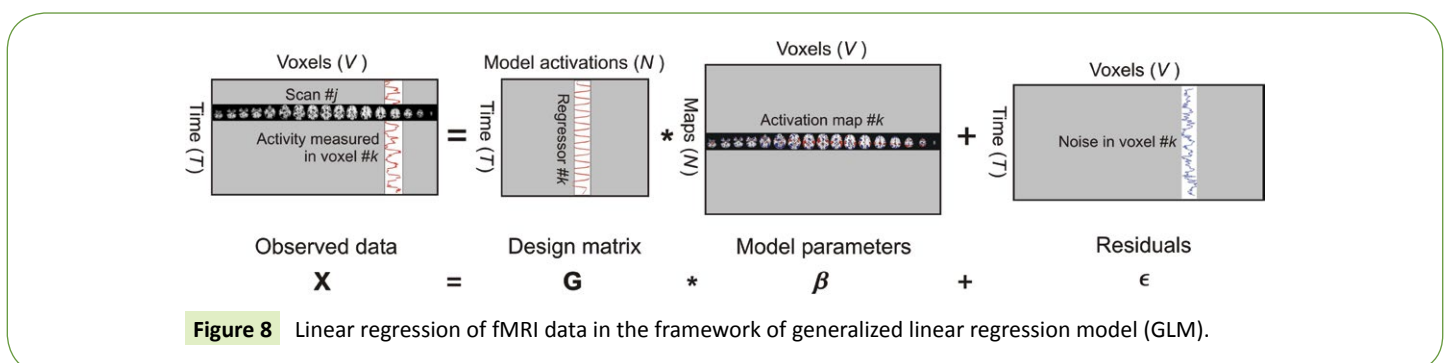
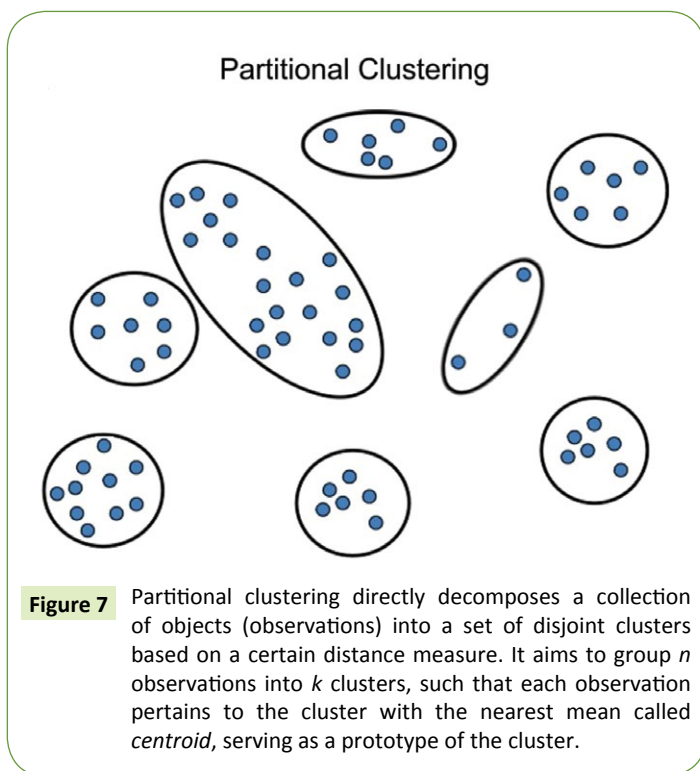
stereotaxic space [42]. A convolution with an isotropic symmetric Gaussian kernel of fixed size is applied prior to statistical analysis. Data preprocessing involves some steps briefly presented hereafter. Spatial smoothing is recommended for:

- (i) match the spatial scale of hemodynamic responses among subjects/sessions;
- (ii) increase the SNR;
- (iii) better data match to the assumptions of the Gaussian Random Field theory (RFT) [41];
- (iv) normalize the error distribution aiming to easier validate inferences based on parametric tests;
- (v) spatial smoothing of temporal autocorrelations to minimize errors in the estimated standard deviation (SD) by increasing the effective degrees of freedom (df) and decreasing its sensitivity to underlying temporal correlation structure [43]. Less spatially variable t -statistic images and lower thresholds ($p < 0.05$) entail better detection of activations and improves the physiological relevance of statistical inference;
- (vi) in multisubject experiments, averaging is necessary to smooth the projected data down to a scale where homologies in functional anatomy are expressed across subjects.

Running more subjects will, nevertheless, improve statistical power much more than moving to higher fields. Apart from its benefits, Gaussian filtering degrades the spatial definition and complicates the statistical analysis since the noise can no longer be considered independent.

The GLM is applied voxel wise on the parameter map resulting from the smoothed data, which does not exploit the spatial correlation between voxels. Hence the observed time series at each voxel are linearly modeled by superimposing a model time course of activation and some residuals like noise and/or measuring errors. If a model of the residuals ϵ exists, then the statistical significance of the regression coefficients and, implicitly, of the modeled hemodynamic changes can be calculated in each voxel via hypothesis testing [41].

Let the matrix $\mathbf{X}_{[T \times V]}$ denote the fMRI data acquired in the experiment, where each matrix element x_{ij} denotes the observed value at time i , $i = 1, 2, \dots, T$ and voxel location j , $j = 1, 2, \dots, V$, V is the total number of voxels in a volume (i.e., full scan) and T is the number time points (i.e., total number of scans in a session).



Then the linear model equates as follows:

$$X = G\beta + \epsilon, \quad (1)$$

where $\mathbf{G}_{[T \times M]}$ is the *design matrix* having the regressors as N column vectors. The row vectors of the matrix $\beta[N \times V]$ are the model parameters of the effects of interest, and the elements in the matrix $\mathbf{s}_{[T \times V]}$ are the residuals of each voxel in each scan. Assuming no temporal correlations in data, a maximum likelihood (ML) estimate for the model parameters β is found by the least-squares method (LSM) as follows:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (2)$$

The parameter selection is carried out by means of a contrast vector, \mathbf{c} , which compares one or multiple parameter values. The multiple hypotheses testing is resolved in statistical parametric mapping (SPM) [6] by considering data as a lattice representation of a continuous Gaussian Random Field (GRF) including the dependencies introduced by the Gaussian spatial smoothing. As such, SPM performs a statistical test on the fitted parameters of the GLM and reveal activation at the spatial locations where the null hypothesis (i.e., no activation) is rejected. The average number of resels (resolution elements containing spatial information) available in the data before smoothing can be considered equal to the number of voxels V . After Gaussian spatial filtering, the average number of resels is reduced [44] down to:

$$\text{resels} = \frac{V}{\prod_{i=1}^3 \text{FWHM}_i}, \quad (3)$$

where FWHM_i is the Gaussian full width at half maximum of the i -th dimension, $i = 1, 2, 3$. The voxel wise test statistics form summary images known as *statistical parametric maps*, which are commonly assessed for statistical significance against the null hypothesis (e.g., no activation). The resulting map of such statistics is a representation of the spatial distribution of functional activity elicited by the experimental paradigm. Less spatially variable t -statistic images and lower thresholds ($p < 0.05$) enforce better detection of activation and improve the physiological relevance of subsequent statistical inference.

References

- Han J, Kamber M, Pei J (2012) Data mining - concepts and techniques (3rd Edn). Elsevier, Morgan Kaufmann, p: 744.
- Cardoso JF (1998) Blind signal separation: statistical principles. Proc. IEEE 86: 2009-2025.
- https://en.wikipedia.org/wiki/Computer_science
- <https://cs.stackexchange.com/>
- https://en.wikipedia.org/wiki/Artificial_intelligence
- Turing A (1950) Computing machinery and intelligence. Mind 59: 433-460.
- Sajeev BU, Thangavel K (2011) Assessment of financial status of SHG members: a clustering approach. Int J Comput Appl 32: 7-15.

During the last two decades fMRI has undergone a quick growing and refining as an interdisciplinary approach by itself and entered firmly in a vast spectrum of scientific disciplines like neuroscience, medical physics, psychology, political science, economics, and law, and so forth. At the same time, brand new approaches have showed up in data collection and processing, experimental design, image reconstruction and enhancement [45-49].

Concluding Remarks

Advanced rsfMRI and DTI techniques have provided means to view the human brain connectome by assessing tract structure, tract connectivity, and functional connectivity. Cortical and white matter damage arising in brain injury can disrupt structural connectivity, thereby affecting patterns of functional brain activity and connectivity. DTI allows to investigate WM damage and to assist in understanding pathological changes in the structural connectivity. rsfMRI permits assessing large-scale functional connectivity and integrity of neuronal networks. These techniques may foster the development of imaging biomarkers of cognitive and neurobehavioral impairments.

Terabyte and petabyte-scale amount of data booming everywhere from business, society, science and engineering, medicine and every aspect of daily life has rendered the traditional data analysis methods unappropriated to efficiently handle unprecedentedly huge data sets. Novel techniques coagulated under the umbrella of DM are called to cope with recent evolution of information complexity aiming to discover new knowledge from newly acquired data and to statistically represent it as prior distributions suitable to validate novel hypotheses. By rapid strides, our society is moving from the data age to the information age with several benefits but some drawbacks as well.

Acknowledgements

The author is grateful to Prof. Stefan Posse (UNM), Prof. Peter Van Zijl (KKI and JHU), Dr. Thomas Balkin (WRAIR) and Dr. Allen Braun (NIH) for their continuous support, genuine data, and stimulating observations and comments whenever asked for.

- Witten IH, Frank E, Hall AM (2011) Data mining: Practical Machine Learning Tools and Techniques (3rd Edn). Burlington, MA 01803, USA: Elsevier Inc.
- Samuel AL (2000) Some studies in machine learning using the game of checkers. IBM J Res Dev 44: 1-2.
- Mitchel TM (1997) Machine Learning. Ithaca NY: McGraw-Hill Science/ Engineering/Math, p: 432.
- Tukey JW (1962) The future of data analysis. Ann Stat 33: 1-67.
- Huber PJ (1994) Huge data sets. In: Compstat- Proceedings in Computational Statistics 11th Symposium (Dutter R, Grossman Edn), pp: 3-13.
- Hadamard J (1923) Lectures on the cauchy's problem in linear partial differential equations. Yale University, CT: Yale University Press, p: 338.

- 14 Iyengar SS, Balakrishnan N (2004) Signal processing for mining information. *IEEE Signal Process Mag* 21: 12-13.
- 15 Beligiannis G, Skarlas L, Likothanassis S (2004) A generic applied evolutionary hybrid technique. *IEEE Signal Process Mag* 21: 28-38.
- 16 Machado A, Gee JC, Campos MFM (2004) Visual data mining for modeling prior distributions in morphometry. *IEEE Signal Process Mag* 21: 20-27.
- 17 Voci F, Eiho S, Sugimoto N, Sekibuchi H (2004) Estimating the gradient threshold in the Perona-Malik equation. *IEEE Signal Process Mag* 21: 39-46.
- 18 Mutihac L, Mutihac R (2008) Mining in chemometrics. *Anal Chim Acta* 612: 1-18.
- 19 Ganapathiraju MK, Klein-Seetharaman J, Balakrishnan N, Reddy R (2004) Characterization of protein secondary structure. *IEEE Signal Process Mag* 21: 78-87.
- 20 Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 24: 417-441.
- 21 Comon P (1994) Independent component analysis, a new concept? *Signal Process* 36: 287-314.
- 22 Goutte C, Toft P, Rostrup E, Nielsen F, Hansen LK (1999) On clustering fMRI time series. *Neuroimage* 9: 298-310.
- 23 Thurstone LL (1931) Multiple factor analysis. *Psychol Rev* 38: 406-427.
- 24 Friedman JH (1987) Exploratory projection pursuit. *J Am Stat Assoc* 82: 249-266.
- 25 Somorjai RL (2002) Exploratory data analysis in functional neuroimaging. *Artif Intell Med* 25: 1-3.
- 26 Mutihac R, Mutihac RC (2007) A comparative study of ICA algorithms for EEG. *Rom Rep Phys* 59: 831-860.
- 27 Windham JP (1985) The use of generalized eigen image analysis for feature selection in MR image sequences. *J Nucl Med* 26: 5.
- 28 <https://pdfs.semanticscholar.org/eafa/ab230c3657603789f3b2bee5ae0a27927c9e.pdf>
- 29 Kohonen T (1995) *Self-Organizing Maps*. New York: Springer, p: 371.
- 30 Mutihac R, Van Hulle MM (2004) Comparison of principal component analysis and independent component analysis for blind source separation. *Rom Rep Phys* 56: 20-32.
- 31 Field DJ (1994) What is the goal of sensory coding? *Neural Comput* 6: 601-191.
- 32 Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? *Network* 3: 213-251.
- 33 Fielding AH (2007) Automated feedback. Faculty of Science LandT conference 2007.
- 34 Paul EB (2012) *Dictionary of algorithms and data structures*. US National Institute of Standards and Technology.
- 35 Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proc* 96: 226-231.
- 36 <https://blogs.sas.com/content/subconsciousmusings/>
- 37 Bezdek JC, Ehrlich R, Full W (1984) FCM: The fuzzy c-means clustering algorithm. *Comput Geosci* 10: 192-203.
- 38 Kassambra A (2017) *Practical guide to cluster analysis in R, unsupervised machine learning (1st Edn)*. STHDA, p: 3
- 39 Buxton RB (2009) *Introduction to functional magnetic resonance imaging: Principles and techniques (2nd Edn)*. New York: Cambridge University Press, p: 22.
- 40 Gregory Ashby F (2011) *Statistical analysis of fMRI data*. Cambridge, Massachusetts: The MIT Press, p: 332.
- 41 Mutihac R (2006) Wavelet-based statistical analysis in functional neuroimaging. *Proceedings of the 6th WSEAS international conference on Wavelet analysis & multirate systems*, pp: 59-77.
- 42 Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers, p: 132.
- 43 Worsley KJ (2005) Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis. *Neuroimage* 26: 635-641.
- 44 Friston KJ, Holmes AP, Worsley KJ, Frackowiak R (1995) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 24: 189-210.
- 45 Mutihac R (2017) *Essentials in brain connectivity*. *J Neurol Clinical Neurosci* 1: 5-11.
- 46 Van De Ville D, Blu T, Unser M (2005) On the multi-dimensional extension of the quincunx subsampling matrix. *IEEE Signal Process Lett* 12: 112-115.
- 47 James CB (1981) *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press, p: 240.
- 48 Somorjai RL, Jarmasz M (2003) Exploratory analysis and data modeling in functional neuroimaging. In: *Exploratory analysis of fMRI data by fuzzy clustering: philosophy, strategy, tactics, implementation (Sommer FT, Wichert A Edn)*. Cambridge, MA: MIT Press, pp: 17-48.
- 49 Parnell LD, Lindenbaum P, Shameer K, Dall'Olio GM, Swan DC, et al. (2011) BioStar: an online question & answer resource for the bioinformatics community. *PLoS Comput Biol* 7: e1002216.