Unsupervised clustering under local constraints of dynamics using multiple equivalence tests

Fuchen Liu [1], Yves Rozenholc [2] & Charles-André Cuenod [3]

[1] *Université Paris Descartes, MAP5 - UMR CNRS 8145 & Intrasense, Paris, fuchen.liu@parisdescartes.fr,*
[2] *Université Paris Descartes, MAP5 - UMR CNRS 8145 & INRIA Saclay Ile-de-France, équipe* Select*, yves.rozenholc@parisdescartes.fr,*
[3] *Laboratoire de Recherche en Imagerie LRI, Université Paris Descartes - INSERM U970 PARCC & Hôpital Européen Georges Pompidou , charles-andre.cuenod@egp.aphp.fr*

Résumé. L'imagerie de perfusion joue un rôle majeur pour tudier la microvascularisation tumorale qui est perturbée par une angiogenèse anormale pendant la croissance de la tumeur. Enregistrant une information dynamique liée à l'injection d'un bolus d'agent de contraste, ce type d'imagerie permet de construire des biomarqueurs diagnostic, prognostic ou de suivi dans le cadre des traitements anti-angiogéniques. Toutefois l'imagerie de perfusion souffre d'un fort niveau de bruit et il est nécessaire d'améliorer le rapport signal sur bruit, par exemple via la construction de régions d'intérêt (ROI) au sein desquelles l'information dynamique est moyennée. Réalisée de façon manuelle ou automatique avec des outils mal adaptés, ces ROI souffrent actuellement d'un manque d'homogénéité ou d'une perte d'information dynamique. Nous proposons de remédier à ces problèmes à travers une classification non supervisée qui préserve les dynamiques et offre un degré d'homogénéité contrôlable. Notre méthode s'appuie sur une utilisation de tests d'équivalence multi-résolution, qui préservent la structure dynamique, et d'un algorithme itératif de type dendrogramme qui protège les propriétés de l'image. La construction itérative s'arrête automatiquement à l'aide d'un contrôle des erreurs de type I et II permettant ainsi de choisir le nombre de classes automatiquement.

Mots-clés. Imagerie de perfusion, test d'équivalence, classification non supervisée, ROI, biomarqueur

Abstract. Perfusion imaging plays an important role in studying tumor microvasculature that undergoes disturbances from abnormal angiogenesis during tumor growth. With the dynamic information related to the injection of a contrast agent bolus, it is possible to build biomarkers for diagnostic, prognosis or treatment monitoring when using anti-angiogenic drugs. However, perfusion imaging suffers from high noise level, hence it is necessary to improve the signal to noise ratio, by constructing for example regions of interest (ROI) in which dynamic information is averaged. Acquired manually or automatically with unsuitable tools, these ROIs are currently suffering from a lack of homogeneity

or a loss of dynamical information. We propose to address these issues through an unsupervised clustering that preserves the dynamics and provides a controllable level of homogeneity. Our method is based on multi-resolution equivalence test, adapted to the dynamic structure, and on a dendrogram-like iterative algorithm that takes into account image specificities. The iterative construction is stopped automatically with the control of Type I and Type II errors permitting an automatic choice of the number of clusters.

Keywords. Perfusion imaging, equivalence test, unsupervised clustering, ROI, biomarker

1 Introduction

In order to improve diagnosis, prognosis and treatment optimization of cancer, more advanced and comprehensive detection and monitoring techniques are required. As tumor growth is associated with a modification of the microvascular function through abnormal angiogenesis, clinicians are therefore keen about monitoring the microvascular function by perfusion imaging techniques, leading to imaging biomarkers. A common perfusion imaging technique is Dynamic Contrast Enhanced (DCE) imaging using Computed Tomography (DCE-CT), Magnetic Resonance Imaging (DCE-MRI) or Ultrasound imaging (DCE-US), which measures tissue enhancement signal at each voxel induced by an intravenous injection of a bolus of contrast agent. By analyzing the information produced in DCE image sequence, either scalar (tissue perfusion, tissue blood volume, mean transit time) or functional parameters (scaled survival function) describing the microvascularization are acquired to build imaging biomarkers.

However, analysis techniques of perfusion imaging suffer from high noise level due to either the need to control X-ray dose or low sensitivity while trying to obtain high resolution. To reduce the noise level, either large manual regions of interests (ROI) are used or the sequence of images are denoised individually by spatial averaging or filtering techniques. However, both of them lose the dynamical information of temporal signal by mixing dynamics which may not be homogeneous. Nevertheless, what we are dealing with in DCE image sequence, after a proper registration, are assumed to be fixed objects with changing contrast. Therefore, with this hypothesis, almost homogeneous objects are expected, which leads us to segment homogeneous clusters from DCE image sequence.

Many works have adapted existing image segmentation methods to DCE-imaging. Thresholding, clustering-based methods such as k-means [Nguyen et al. (2014)], fuzzy c-means [Chen et al. (2006)] as well as support vector machine [Torheim et al. (2014)] and artificial neural network [Szabó et al. (2004)] have been adopted in conjunction with a dimension-reduced feature space, such as qualitative [Lavini et al. (2006)] or kinetic [Chen et al. (2011)] parameters and sub-space produced by principal component analysis, spectral embedding [Agner et al. (2013)] and wavelets techniques [Li et al. (2012)], instead of original data space, in order to reduce the computational complexity and noise. However, a fair amount of dynamical information suffers a loss due to this dimension reduction.

Besides, in order to achieve the automatic selection of the number of clusters by using an information criterion as in Tartare et al. (2014), same method has to be performed repeatedly for each given number of clusters, which massively increases the computational complexity specially for large target ROI with maybe over one hundred clusters.

To overcome these issues, a couple of methods have been recently developed by Rozenholc et al. (2010) and Mohajer et al. (2012). The latter one is a two-resolution level method based on hierarchical clustering, using an Euclidean distance adapted to the characteristics of DCE signal curves. However, due to the use of two-resolution levels, the resulting clusters suffer from unexpected shapes, moreover the cluster homogeneity is not satisfying. On the other hand, DynClust by Yves et al. (2010) is a two step, point-wise denoising and clustering, method based on the growth of homogeneous regions and adaptive multiple test. Automatic selection of the number of clusters is ensured by the user-defined test level. The neighborhood construction during the denoising step ensures that large organs, tumors, metastasis and vessels are well clustered. However, the denoising step is time-consuming and the choice of the alternative hypothesis is questionable for the purpose of comparison. The latter causes a lack of mathematical basis for this method.

We propose a new method for clustering dynamical images, relying on multiple equivalence test, which takes care of local homogeneity and can also merge similar disconnected features. The optimal number of clusters is selected automatically through a proper control of the type I and type II errors.

2 Statistical model

We consider a DCE imaging sequence, that we will also call dynamical image, consisting of a finite sequence of noisy images indexed by both time and space:

$$I = \{I^x := (I^x(t_1), \dots, I^x(t_N)), x \in \mathcal{X}, t_1 \leqslant t_2 \leqslant \dots \leqslant t_N\},$$

where $I^x(t_j)$ denotes the noisy enhancement at the j-th acquisition time and voxel location x of the finite voxel grid $\mathcal{X}$. We assume that the j-th observable gray level $I^x(t_j)$ may be written as

$$I^x(t_j) = i^x(t_j) + \sigma\varepsilon_j^x,$$

where $i^x(t_j)$ denotes the true but unobservable gray level, ε_j^x denotes a standardized noise and σ the noise level. We assume that the noises ε_j^x are independent with respect to both space location x and time index j. To help the presentation, we will assume that σ is known and $\sigma = 1$. In practice, the knowledge of σ may be either ensured by a proper calibration or replaced by a consistent estimation.

Our method is based on the statistical comparison, for two sets of voxels X and Y, between the two dynamics estimated by averaging: $\bar{I}^X = \sum_{x\in X} I^x/|X|$, $\bar{I}^Y = \sum_{y\in Y} I^y/|Y|$. Assuming $X \cap Y = \emptyset$ such that

$$D^{XY} = \bar{I}^X - \bar{I}^Y \sim \mathcal{N}(d^{XY}, \Sigma_N),$$

where $d^{XY} = \sum_{x \in X} i^x/|X| - \sum_{y \in Y} i^y/|Y|$ and $\Sigma_N = (1/|X| + 1/|Y|)\,\mathrm{Id}_N$, the two sets are considered indistinguishable if d^{XY} does not deviate significantly from the zero vector. More precisely, we aim to build a so-called "equivalence" test

$$\mathcal{H}_0 : d^{XY} \neq 0 \quad \text{v.s.} \quad \mathcal{H}_1 : d^{XY} = 0$$

and decide that X and Y are time homogeneous when $\mathcal{H}_1$ is true.

3 Equivalence testing

In hypotheses testing, the research or alternative hypothesis represents what the study aims to put in evidence. The burden of proof is on the alternative in the sense that it is established only if there is enough evidence in its favor, which in conventional (two-sided) comparative studies, is the hypothesis of difference. In contrast, the goal of equivalence test [Walker and Nowacki (2010)] is to demonstrate equivalence. Therefore the burden of proof rests on equivalence. In essence, the null and alternative hypotheses in equivalence test are simply those of a conventional comparative study reversed. The term "equivalence" means that two random variables are close enough to be distinguishable.

The construction of equivalence test under Gaussian assumption can be found in Wellek (2010). Using such equivalence test in our model, we would have to run N comparisons (one per coordinates) and would face a multiplicity problem. To control this multiplicity, we follow the works of Baraud et al. (2005) and Durot and Rozenholc (2006), and use a dyadic decomposition of the time index to project the dynamics onto sub-spaces resulting that only $K_{\max} := \lfloor \log_2 N \rfloor$ tests are needed to perform the comparison. For $K = 1, \ldots, K_{\max}$, we denote by Π_K the projection in $\mathbb{R}^N$ onto the piecewise constant vectors on the dyadic partition made of 2^K (almost) regular intervals and consider the equivalence test

$$\mathcal{H}_0^K : \| \Pi_K(d^{XY}) \|^2 \neq 0 \quad \text{v.s.} \quad \mathcal{H}_1^K : \| \Pi_K(d^{XY}) \|^2 = 0.$$

We will say that $\mathcal{H}_1$ is true if and only if all $\mathcal{H}_1^K$ are true, that is

$$\mathcal{H}_0 = \bigcup_{K=1}^{K_{\max}} \mathcal{H}_0^K \quad \text{v.s.} \quad \mathcal{H}_1 = \bigcap_{K=1}^{K_{\max}} \mathcal{H}_1^K,$$

which is well known as an intersection-union test (IUT) from Berger and Hsu (1996). In this setting the following result holds

Theorem 1 (Berger and Hsu (1996), page 6) *Let R_K denote a rejection region for a test of $\mathcal{H}_0^K$ at level α, then the IUT with rejection region $R = \bigcap_{K=1}^{K_{max}} R_K$ is of level α.*

Corollary 1 *If p (resp. p_K) is the p-value for $\mathcal{H}_0$ (resp. $\mathcal{H}_0^K$) then $p = \max_K(p_K)$.*

To formalize the construction of the K-th equivalence test, we introduce the equivalence margin Δ, and compute the K-th p-value as $P(\chi^2(2^K, \Delta) \leqslant \| \Pi_K(D^{XY}) \|^2)$, where $\chi^2(2^K, \Delta)$ denotes a Δ-shifted χ^2 with 2^K degrees of freedom.

4 Dendogram-like clustering using equivalence test

A general setup of clustering problem is to produce a partition with n clusters

$$\mathcal{X} = C_1 \cup \ldots \cup C_n, \quad C_s \cap C_{s'} = \emptyset, \quad 1 \leqslant s, s' \leqslant n.$$

Using equivalence test, we aim to build this partition to be minimal with respect to n and to satisfy that $\forall x \in C_s$ and $\forall y \in C_{s'}$, $\mathcal{H}_1$ is true if $s \neq s'$ and $\mathcal{H}_0$ is true otherwise. The computational complexity of this program is unfortunately $O(2^{|\mathcal{X}|})$, which is impossible when $|\mathcal{X}|$ is already one hundred. Moreover, we would like to take into account local properties of DCE image sequence such as local homogeneity.

We adopt a bottom-top dendrogram construction by building on a pseudo-dissimilarity derived from the p-value of our equivalence test for spacial neighbors and which is infinite for non spacial neighbors. By controlling the type I and type II errors, we compute the equivalence margin Δ and provide an automatic way to stop our construction, which defines both the number of clusters and the partition. This local strategy can be followed by a global one, which aims at recovering disconnected clusters with equivalent dynamics. To this end, starting from the previous partition, we adapt the previous construction without consideration to the neighborhood structure while computing the pseudo-dissimilarity.

The complexity of the local procedure is controlled by the spatial localization while those of the global procedure is controlled thanks to the expected small size of the partition obtained at the first step.

References

[1] Nguyen H. T., Jia G., Shah Z. K., Pohar K., Mortazavi A., Zynger D. L., Wei L., Yang X., Clark D. and Knopp M. V. (2014), Prediction of chemotherapeutic response in bladder cancer using K-means clustering of dynamic contrast-enhanced (DCE)-MRI pharmacokinetic parameters, *Journal Of Magnetic Resonance Imaging*.

[2] Chen W., Giger M. L., Bick U. and Newstead G. M. (2006), Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI, *Medical Physics*, 33(8), 2878–2887.

[3] Torheim, T., Malinen, E., Kvaal, K., Lyng, H., Indahl, U. G., Andersen, E. K. F., Futsaether, C. M. (2014), Classification of Dynamic Contrast Enhanced MR Images of Cervical Cancers Using Texture Analysis and Support Vector Machines, *IEEE Transactions on Medical Imaging*, 33, 1648–1656.

[4] Szabó B. K., Aspelin P. and Wiberg M. K. (2004), Neural network approach to the segmentation and classification of dynamic magnetic resonance images of the breast: Comparison with empiric and quantitative kinetic parameters, *Academic Radiology*, 11(12), 1344-1354.

[5] Lavini C., de Jonge M. C., van de Sande M. G., Tak P. P., Nederveen A. J. and Maas M. (2006), Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its

application to the imaging of the musculoskeletal system, *Magnetic Resonance Imaging*, 25, 604–612.
[6] Chen L., Choyke P. L., Chan T., Chi C., Wang G. and Wang Y. (2011), Tissue-Specific Compartmental Analysis for Dynamic Contrast-Enhanced MR Imaging of Complex Tumors, *IEEE Transactions on medical imaging*, 30(12).
[7] Agner S. C., Xu J. and Madabhushi A. (2013), Spectral embedding based active contour (SEAC) for lesion segmentation on breast dynamic contrast enhanced magnetic resonance imaging, *Medical Physics*, 40(3).
[8] Li S., Zöllner F. G., Merrem A. D., Peng Y., Roervik J., Lundervold A. and Schad L. R. (2012), Wavelet-based segmentation of renal compartments in DCE-MRI of human kidney: Initial results in patients and healthy volunteers, *Computerized Medical Imaging and Graphics*, 36, 108–118.
[9] Tartare G., Hamad D., Azahaf M., Puech P. and Betrouni N. (2014), Spectral clustering applied for dynamic contrast-enhanced MR analysis of time-intensity curves, *Computerized Medical Imaging and Graphics*.
[10] Rozenholc Y., Reißd M., Balvayc D. and Cuénod C. A. (2010), Growing time-homogeneous neighborhoods for denoising and clustering Dynamic Contrast Enhanced-CT sequences.
[11] Mohajer M., Schmid V. J., Engels N. A., Noel P. B., Rummeny E. and Englmeier K. H. (2012), Stepwise heterogeneity analysis of breast tumors in perfusion DCE-MRI datasets, *Medical Imaging 2012: Biomedical Applications in Molecular, Structural, and Functional Imaging. Proc. SPIE 8317*.
[12] Walker E. and Nowacki A. S. (2010), Understanding Equivalence and Noninferiority Testing, *Journal of General Internal Medicine*, 26(2), 192–196.
[13] Wellek S. (2010), *Testing Statistical Hypotheses of Equivalence and Noninferiority, Second Edition*, Chapman and Hall/CRC.
[14] Baraud Y., Huet S. and Laurent B. (2003), Adaptive tests of linear hypotheses by model selection, *The Annals of Statistics*, 31(1), 225–251.
[15] Baraud Y., Huet S. and Laurent B. (2005), Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function, *The Annals of Statistics*, 33(1), 214–257.
[16] Durot C. and Rozenholc Y.(2006), An adaptive test for zero mean, *Mathematical Methods of Statistics*, 15(1), 26-60.
[17] Berger R. L. and Hsu J. C. (1996), Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets, *Statistical Science*, 11(4), 283–319.