Topological Radiomics (TOPiomics): Early Detection of Genetic Abnormalities in Cancer Treatment Evolution

Debora Gil, Oriol Ramos, and Raquel Perez

Abstract Abnormalities in radiomic measures correlate to genomic alterations prone to alter the outcome of personalized anti-cancer treatments. TOPiomics is a new method for the early detection of variations in tumor imaging phenotype from a topological structure in multi-view radiomic spaces.

AQ1

1 Introduction

In the era of precision medicine, cancer therapies are tailored to the specific genetic makeup of a tumour. A main challenge during treatment is the early detection of variations in tumour phenotype that might alter the expected outcome. Radiomics [1] is an emerging area that converts medical imaging data into large amount of multiview measures (imaging phenotype) of the whole tumour correlated with genomics. Although abnormal radiomic features could be predictive early response biomarkers to cancer treatments, there are no methods specifically developed for detection of abnormalities (outliers). There are two main types of outliers in radiomic multi-view spaces [2]. Samples with inconsistent features with respect their class population (class outliers associated to a change in the mutation type) and samples with abnormal feature values not expected for any of the classes (attribute outlier associated to new unseen mutations).

Detection of abnormal radiomic features should model multi-view spaces with Small Sample Size (SSS) data prone to have a complex manifold structure. A main

D. Gil (✉) · O. Ramos
Computer Vision Center and Computer Science Department, Universitat Autonoma de Barcelona, Barcelona, Spain
e-mail: debora@cvc.uab.es

O. Ramos
e-mail: oriolrt@cvc.uab.es

R. Perez
Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain
e-mail: rperez@vhio.net

M. Alberich-Carramiñana et al. (eds.), *Extended Abstracts GEOMVAP 2019*,
Trends in Mathematics 15,
https://doi.org/10.1007/978-3-030-84800-2_15

516361_1_En_15_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:26/8/2021 Pages: xxx Layout: T1-Standard

pitfall in current state of the art is the use of generic machine learning and statistical tools borrowed from other fields of application which fall short under the specific requirements of radiomics [3].

Existing methods for detection of outliers can be categorized into global approaches and local approaches. Global methods are population based and model the distribution in the feature space of a set of (annotated) samples. Global approaches are bad posed in the case of SSS unbalanced problems, which are common in many application areas like clinical decision support systems or personalized models. Local methods are based on a description of the structure of each sample's neighbors in the feature space. These description is used to compute measures of outlierness. A delicate requirement is the definition of sample's neighborhoods, which is mostly based on Euclidean distances. Such an approach can fail in the case of SSS problems in high dimensional spaces, which are prone to be arranged as a topological manifold.

The goal of TOPiomics is the early detection of variations in tumour imaging phenotype using a topological signature of abnormality obtained from the topological structure of SSS data in multi-view radiomic spaces.

2 Methods

TOPiomics is a local approach based on the communities (group of nodes with a given specific connectivity) of a graph encoding the structure of radiomics feature space. Features are given by quantities extracted from medical scans prone to correlate to treatment outcome, referred to as label. In the context of radiomics multimodal representations, there are two types of outliers: attribute outliers and class outliers. Attribute outliers are samples with abnormal feature values not expected for any of the classes, while class outliers are samples labelled differently across views.

Figure 1 sketches the main steps of TOPiomics. First, for each radiomic view (like the one shown in Fig. 1a), we encode the local structure of samples using the graph representing their mutual k-nearest neighbor (Fig. 1b). Second, we use methods for dynamical analysis of social networks to compute the graph communities (Fig. 1c) that define a set of neighborhoods. Isolated nodes not belonging to any community are attribute outliers, while class outliers should belong to communities with an heterogeneous distribution of labels. Finally, we define a local measure of abnormality from several probabilistic measures (Fig. 1d) of each sample heterogeneity computed in its set of neighborhoods.

The graph is given by the adjacency matrix of the mutual k-nearest neighbor of a set of samples. Let $D := \{(\mathbf{V}^i, \ell_{\mathbf{V}^i}) | \mathbf{V}^i = (v_1^i, \ldots, v_n^i) \in \mathbb{R}^n, \ell_{\mathbf{V}^i} \in \{1, \ldots, n_l\}\}_{i=1}^N$ be a set of N labelled points in an n-dimensional feature space endowed with a distance, namely d. For any positive integer, k, let $\mathrm{kNN}(\mathbf{V}^i)$ denote the set of $\mathbf{V}^i$ k-nearest neighbors. Then, the graph connectivity is given by the following adjacency matrix:

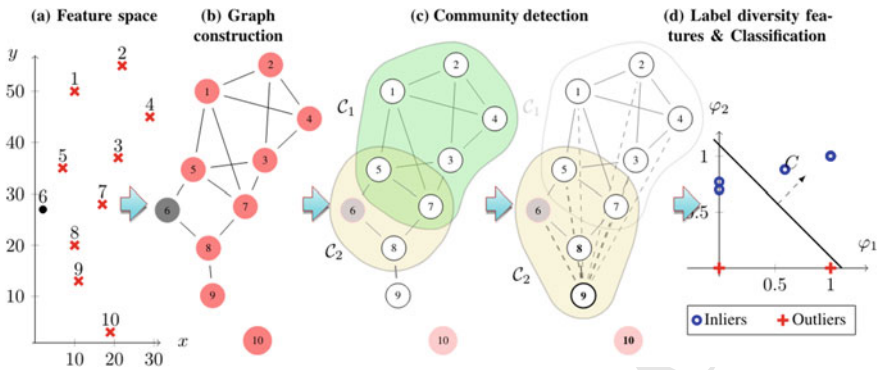

Fig. 1 TOPiomics workflow

$$a(\mathbf{V}_i, \mathbf{V}_j) = \begin{cases} \frac{1}{d(\mathbf{V}_i, \mathbf{V}_j)+1} & \text{if } \mathbf{V}_j \in \text{kNN}(\mathbf{V}_i) \text{ and } \mathbf{V}_i \in \text{kNN}(\mathbf{V}_j) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

for $d(\mathbf{V}_i, \mathbf{V}_j)$ the distance between $\mathbf{V}_j$ and $\mathbf{V}_i$.

In order to alleviate the impact of the parameters (the number of neighbors in this case) involved in the computation of (1), communities are computed using criteria for dynamic computation of communities [4] to extend an initial set of communities. The initial communities are given by Percolation clusters [5] which are defined as maximal unions of adjacent k-cliques (fully connected subgraphs of order k sharing (k-1)-nodes). Percolation communities are prone to exclude many points that are not actual attribute outliers [6]. An isolated node, $\mathbf{W}$, is added to an initial community, $\mathcal{C}$, if it fulfills that:

$$CS(\mathcal{C}, \mathbf{W}) \geq \delta IC(\mathcal{C}) \tag{2}$$

for $\delta \in [0, 1]$ a tolerance parameter, $IC(\mathcal{C})$ a measure of the community internal connectivity and $CS(\mathcal{C}, \mathbf{W})$ a measure of the connectivity between $\mathbf{W}$ and the community $\mathcal{C}$. Both measures are computed from a function of the degree of the community nodes as follows.

Let $G^{\mathcal{C}}$ be the subgraph induced by $\mathcal{C}$ and G^{σ} the subgraph induced by all nodes that belong to the set, namely σ, of initial communities. Then, for all $\mathbf{V} \in \mathcal{C}$ we can define the following function, $\rho_{\mathcal{C}}(\mathbf{V})$, measuring its belongingness to the community:

$$\rho_{\mathcal{C}}(\mathbf{V}) := \frac{\deg^{\mathcal{C}}(\mathbf{V})}{\deg^{\sigma}(\mathbf{V})} \tag{3}$$

being $\deg^{\mathcal{C}}(\mathbf{V})$ the degree of $\mathbf{V}$ in $G^{\mathcal{C}}$ and $\deg^{\sigma}(\mathbf{V})$ the degree of $\mathbf{V}$ in G^{σ}. The measure of $\mathcal{C}$ internal connectivity is defined from $\rho_{\mathcal{C}}(\mathbf{V})$ as:

$$IC(\mathcal{C}) := \sum_{\mathbf{V} \in \mathcal{C}} \rho_{\mathcal{C}}(\mathbf{V}) \tag{4}$$

Table 1 Assessment of Performance

DataSet	Method	Outlier configuration		
		2-8	5-5	8-2
Iris	HOAD	0.167 ± 0.057	0.309 ± 0.063	0.430 ± 0.055
	DMOD	0.909 ± 0.044	0.831 ± 0.038	0.799 ± 0.068
	TOPiomics	0.975 ± 0.024	0.971 ± 0.023	0.97 ± 0.021
Breast	HOAD	0.538 ± 0.027	0.597 ± 0.038	0.643 ± 0.008
	DMOD	0.657 ± 0.017	0.720 ± 0.013	0.799 ± 0.016
	TOPiomics	0.838 ± 0.022	0.897 ± 0.020	0.91 ± 0.014
Ionosphere	HOAD	0.489 ± 0.079	0.477 ± 0.072	0.444 ± 0.065
	DMOD	0.818 ± 0.018	0.787 ± 0.039	0.784 ± 0.037
	TOPiomics	0.854 ± 0.019	0.827 ± 0.025	0.791 ± 0.036

The measure of the connectivity between $\mathbf{W}$ and $\mathcal{C}$ is defined from $\rho_{\mathcal{C}}(\mathbf{V})$ as:

$$CS(\mathcal{C}, \mathbf{W}) := \sum_{\mathbf{V} \in \mathcal{C}} \rho_{\mathcal{C}}(\mathbf{V}) \frac{1}{d(\mathbf{W}, \mathbf{V}) + 1} = \sum_{\mathbf{V} \in \mathcal{C}} \rho_{\mathcal{C}}(\mathbf{V}) a(\mathbf{W}, \mathbf{V}) \tag{5}$$

For the final measure of outlierness, we define a 2-dimensional feature space given by functions of node label entropy and probability in the communities it belongs to. Functions are normalized in [0, 1] in such a way that inliers correspond to values around (1, 1). A classifier provides our final score of outlier-ness.

3 Experiments

TOPiomics performance has been assessed in UCI[1] datasets altered to have different % of attribute and class outliers. We have followed the experimental settings described in [2]. In particular, we considered 3 combinations of percentages in attribute and class outliers ({(8%, 2%), (5%, 5%), (2%, 8%)}) and a multi-view setting. For each outlier configuration, we repeated the experiment 30 times for statistical analysis of results. TOPiomics has been compared to the state-of-art methods reported in [2] in terms of Area Under the ROC Curve (AUC).

Table 1 reports a statistical summary (average $\pm$ standard deviation) for the results obtained for TOPiomics, HOAD [7] and DMOD [2] in Iris (2-views), Breast (3-views) and Ionosphere (3-views) UCI datasets. Ranges indicate that TOPiomics is a better performance regardless database and outlier configuration.

[1] https://archive.ics.uci.edu/ml/datasets.php.

4 Conclusions

TOPiomics description is able to model the complex structure of radiomics SSS multi-view data. Its non-parametric local description endows TOPiomics with high robustness to detect abnormalities in SSS contexts, while its view-sensitive approach allows early detection of abnormal imaging phenotypes. Therefore, TOPiomics could be a unique specific technique to define robust imaging biomarkers for outcome in cancer treatment follow-up that will improve cancer patients care by optimizing treatment selection and sequence.

Acknowledgements This project has received funding from the ATTRACT project funded by the EC under Grant Agreement 777222. The work has also been partially funded by Spanish Projects FIS-G64384969, RTI2018-095209-B-C21, RTI2018-095645-B-C21, Generalitat de Catalunya 2017-SGR-1624, 2017 SGR 1783 and CERCA-Programme. The Titan V and Titan X Pascal used for this research was donated by the NVIDIA Corporation. R.Perez is supported by a PCF-Young Investigator Award and Instituto de Salud Carlos III-Investigacin en Salud (PI18/01395). DGil is a Serra Hunter Fellow.

References

1. P. Lambin et al., Radiomics: the bridge between medical imaging and personalized medicine. Nat. Rev. **12**, 749–53 (2017)
2. H. Zhao et al., Consensus regularized multi-view outlier detection. IEEE Trans. Imag. Proc. **27**(1), 1 (2018)
3. JP. Cohen et al., Distribution matching losses can hallucinate features in medical image translation, MICCAI 2018
4. R. Cazabet et al., Detection of overlapping communities in dynamical social networks, Social Comp (2010)
5. C. Domb, E. Stoll, T. Schneider, Percolation clusters. Contemp. Phys. **21**(6), 577–592 (1980)
6. J Mielgo, Analysis of Community Detection Algorithms for Image Annotation, MathMods Master Thesis (2017)
7. J. Gao et al., A multi-graph spectral framework for mining multi-source anomalies, Graph Embedding for Pattern Analysis, pp. 205–228 (2013)
8. D. Gil et al., Classification of Confocal Endomicroscopy Patterns for Diagnosis of Lung Cancer, CLIP 2017

Author Queries

Chapter 15

Query Refs.	Details Required	Author's response
AQ1	Please confirm if the corresponding author is correctly identified. Amend if necessary.	
AQ2	Reference [8] is given in the list but not cited in the text. Please cite them in text or delete them from the list.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	··· under matter to remain	New matter followed by
Insert in text the matter indicated in the margin		or (x2)
Delete	through single character, rule or underline or through all characters to be deleted	or (x2)
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	