

STATISTICAL MODELING AND ANALYSIS OF RADIO-INDUCED ADVERSE EFFECTS BASED ON IN VITRO AND IN VIVO DATA

ABSTRACT

In this work we address the problem of adverse effects induced by radiotherapy on healthy tissues. The goal is to propose a mathematical framework to compare the effects of different irradiation modalities, to be able to ultimately choose those treatments that produce the minimal amounts of adverse effects for potential use in the clinical setting. The adverse effects are studied in the context of two types of data: in terms of the in vitro omic response of human endothelial cells, and in terms of the adverse effects observed on mice in the framework of in vivo experiments. In the in vitro setting, we encounter the problem of extracting key information from complex temporal data that cannot be treated with the methods available in literature. We model the fold changes, the object that encodes the difference in the effect of two experimental conditions, in the way that allows to take into account the uncertainties of measurements as well as the correlations between the observed entities. We construct a distance, with a further generalization to a dissimilarity measure, allowing to compare the fold changes in terms of all the important statistical properties. Finally, we propose a computationally efficient algorithm performing clustering jointly with temporal alignment of the fold changes. The key features extracted through the latter are visualized using two types of network representations, for the purpose of facilitating biological interpretation. In the in vivo setting, the statistical challenge is to establish a predictive link between variables that, due to the specificities of the experimental design, can never be observed on the same animals. In the context of not having access to joint distributions, we leverage the additional information on the observed groups to infer the linear regression model. We propose two estimators of the regression parameters, one based on the method of moments and the other based on optimal transport, as well as the estimators for the confidence intervals based on the stratified bootstrap procedure.

Key words: Radiotherapy, Complex temporal data, Joint clustering with alignment, Omic network inference, Data fusion, Wasserstein distance.

RÉSUMÉ

Dans ce travail nous abordons le problème des effets indésirables induits par la radiothérapie sur les tissus sains. L'objectif est de proposer un cadre mathématique pour comparer les effets de différentes modalités d'irradiation, afin de pouvoir éventuellement choisir les traitements qui produisent le moins d'effets indésirables pour l'utilisation potentielle en clinique. Les effets secondaires sont étudiés dans le cadre de deux types de données : en termes de réponse omique *in vitro* des cellules endothéliales humaines, et en termes d'effets indésirables observés sur des souris dans le cadre d'expérimentations *in vivo*. Dans le cadre *in vitro*, nous rencontrons le problème de l'extraction d'informations clés à partir de données temporelles complexes qui ne peuvent pas être traitées avec les méthodes disponibles dans la littérature. Nous modélisons le fold change radio-induit, l'objet qui code la différence d'effet de deux conditions expérimentales, d'une manière qui permet de prendre en compte les incertitudes des mesures ainsi que les corrélations entre les entités observées. Nous construisons une distance, avec une généralisation ultérieure à une mesure de dissimilarité, permettant de comparer les fold changes en termes de toutes leurs propriétés statistiques importantes. Enfin, nous proposons un algorithme computationnellement efficace effectuant le clustering joint avec l'alignement temporel des fold changes. Les caractéristiques clés extraites de ces dernières sont visualisées à l'aide de deux types de représentations de réseau, dans le but de faciliter l'interprétation biologique. Dans le cadre *in vivo*, l'enjeu statistique est d'établir un lien prédictif entre des variables qui, en raison des spécificités du design expérimental, ne pourront jamais être observées sur les mêmes animaux. Dans le contexte de ne pas avoir accès aux lois jointes, nous exploitons les informations supplémentaires sur les groupes observés pour déduire le modèle de régression linéaire. Nous proposons deux estimateurs des paramètres de régression, l'un basé sur la méthode des moments et l'autre basé sur le transport optimal, ainsi que des estimateurs des intervalles de confiance basés sur le bootstrap stratifié.

Mots clés : Radiothérapie, Données temporelles complexes, Clustering joint avec alignement, Inférence de réseau omic, Fusion de données, Distance de Wasserstein.