



Robust feature selection to predict tumor treatment outcome



Hongmei Mi^{a,*}, Caroline Petitjean^a, Bernard Dubray^{a,b}, Pierre Vera^{a,b}, Su Ruan^a

^a QUANTification en Imagerie Fonctionnelle – Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes (EA4108 – FR CNRS 3638), University of Rouen, 22, Boulevard GAMBETTA, 76183 Rouen, France

^b Centre Henri Becquerel, Rue d'Amiens, 76038 Rouen, France

ARTICLE INFO

Article history:

Received 6 January 2015

Received in revised form 1 July 2015

Accepted 1 July 2015

Keywords:

Hierarchical forward feature selection

Prior knowledge

Prediction

Small sample

Positron emission tomography

Support vector machine

ABSTRACT

Objective: Recurrence of cancer after treatment increases the risk of death. The ability to predict the treatment outcome can help to design the treatment planning and can thus be beneficial to the patient. We aim to select predictive features from clinical and PET (positron emission tomography) based features, in order to provide doctors with informative factors so as to anticipate the outcome of the patient treatment. **Methods:** In order to overcome the small sample size problem of datasets usually met in the medical domain, we propose a novel wrapper feature selection algorithm, named HFS (hierarchical forward selection), which searches forward in a hierarchical feature subset space. Feature subsets are iteratively evaluated with the prediction performance using SVM (support vector machine). All feature subsets performing better than those at the preceding iteration are retained. Moreover, as SUV (standardized uptake value) based features have been recognized as significant predictive factors for a patient outcome, we propose to incorporate this prior knowledge into the selection procedure to improve its robustness and reduce its computational cost.

Results: Two real-world datasets from cancer patients are included in the evaluation. We extract dozens of clinical and PET-based features to characterize the patient's state, including SUV parameters and texture features. We use leave-one-out cross-validation to evaluate the prediction performance, in terms of prediction accuracy and robustness. Using SVM as the classifier, our HFS method produces accuracy values of 100% and 94% on the two datasets, respectively, and robustness values of 89% and 96%. Without accuracy loss, the prior-based version (pHFS) improves the robustness up to 100% and 98% on the two datasets, respectively.

Conclusions: Compared with other feature selection methods, the proposed HFS and pHFS provide the most promising results. For our HFS method, we have empirically shown that the addition of prior knowledge improves the robustness and accelerates the convergence.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

A proportion of cancer patients develop post-treatment recurrence, which increases the risk of death. The ability to predict the treatment outcome prior to or even during the treatment can be of clinical value, as in this case therapy could be individually tailored according to the prediction [1]. Imaging plays a crucial role as it allows for a non-invasive following up of the tumor response to the treatment. Indeed, functional information gathered by positron emission tomography (PET) using the radiotracer FDG has already shown its predictive value for tumor response to treatment in several cancers, including esophageal [2], lung [3] and cervix [4,5]. Well-explored FDG-PET imaging features include, but are not

limited to, metabolic tumor volume (MTV) and total lesion glycolysis (TLG), as functional indices describing metabolic tumor burden, and standardized uptake values (SUVs) describing FDG uptake within a region of interest (ROI), e.g. SUVmean, SUVpeak, or single pixel (SUVmax) [2,4]. Characterization of PET images through texture analysis [6,7], tumor shape [6] and intensity volume histogram [6] may also have potential predictive value for treatment outcome, providing additional and complementary indices. The values of these parameters before treatment, as well as during treatment, are claimed as predictors for recurrence [5,8]. The knowledge of the most predictive factors for treatment outcome is valuable, as doctors can then make personalized treatment plan. However, there is no clear consensus regarding the optimum predictive factors. Feature selection techniques facilitate the interpretation of data by identifying meaningful features for patient outcome prediction. However, feature selection methods are hardly robust to small sized datasets of the order of a few dozen samples, as often in the

* Corresponding author.

E-mail address: hongmei_mi@hotmail.com (H. Mi).

medical domain. In this study, we intend to design a reliable feature selection algorithm to reduce the small size effect, in order to find discriminatory features among a number of patient' clinical and PET-based features, that allow to accurately predict the treatment outcome.

Feature selection methods can be broadly classified into *filter*, *embedded* and *wrapper* methods [9]. Filter methods consist in applying a statistical measure to assign a score to each feature, independently of a classifier. RELIEF (RElevance In Estimating Features, [10,11]) is considered as one of the most successful filter algorithms, where a margin-based criterion is used to rank the features. Authors in [12] propose the FAST (Feature Assessment by Sliding Thresholds) method, based on the area under the receiver operating characteristic which is generated by sliding threshold values in one dimensional feature space. However, these univariate ranking methods cannot take the interaction between features into account. Studies in [13] have pointed out that features which are irrelevant on their own can be useful in conjunction with other features, and that the combination of two highly correlated features can be more useful than each feature independently. Feature subset selection methods evaluate subsets of features together, as opposed to ranking features according to their individual discrimination power. Feature selection with kernel class separability (KCS) ranks feature subsets according to their class separability. It is claimed robust to small size samples and to the presence of noisy features [14,15]. However, as for other feature ranking methods, a threshold value or a number of features has to be specified by the user to obtain the final subset.

Embedded methods include feature selection as part of the training process, and are usually specific to given learning machines. For example, CART has a built-in mechanism to perform feature selection [16]. To split a node into two, one feature chosen by certain rules is used to differentiate observations, which also raises the univariate problem. Guyon et al. propose a method using the support vector machine (SVM) [17] based on Recursive Feature Elimination, named SVMRFE [18]. Starting with the whole feature space, SVMRFE progressively eliminates the least relevant ones, whose removal minimizes the variations of feature weights, until a user-defined number of features remains. The SVMRFE method yields nested feature subsets and still retains the risk of removing useful features as complementary to others. The SVMRFE method has been used in [19] to select predictive features from a number of 167 features composed of patient's clinical, demographic and imaging data, for the prediction of the treatment response in 20 patients with esophageal cancer. The best 10-fold cross-validation prediction is obtained using SVM with 17 selected features.

The principle of wrapper methods is to compare different feature combinations based on the classification performance. Forward selections perform well when the optimal subset is small. In the literature, sequential forward selection (SFS [20]) and sequential forward floating selection (SFFS [21]) are two representative algorithms. Starting from an empty set, SFS repeatedly selects the best feature among the remaining features, and adds it into the set of already selected ones [22]. The SFS method has been used in [6] to select relevant features among 19 PET imaging features for the outcome prediction, on two datasets composed of 14 cervix cancer patients and 9 head and neck cancer patients, respectively. Logistic regression models of the order of two are then constructed. The SFS method yields a nested feature subset, since once a feature is included, it cannot be reconsidered afterwards. This increases the risk of being trapped in suboptimal solutions. To solve the nesting problem, SFFS performs exclusion steps after each inclusion step [23], improving the chance to find the most relevant features. However, for small sample sized datasets, there is a high probability that some feature subsets will produce similar values or even the same value as the evaluation measure. This fact

increases the possibility for SFS and SFFS to be trapped in local minima.

To reduce the problem caused by small sample size in existing wrappers, we propose a novel wrapper forward selection algorithm where SVM is used as the classifier and a feature subset may be retained or discarded based on its prediction accuracy. The novelty of the proposed wrapper is its search strategy. As different feature subsets may yield similar or identical accuracy values on small sample datasets, we propose to keep all candidate feature subsets that increase the accuracy at an iteration, with respect to the previous iteration. Each retained feature subset is then used to generate its successors by adding one feature from the original feature pool which is not yet included in this feature subset. These successors are evaluated at the following iteration. The forward inclusion step does not stop until the accuracy value stops increasing. As a result, the search procedure examines a hierarchical binary search space. Compared with other wrappers where the inclusion or exclusion of a feature is operated on a single feature subset with the highest accuracy at an iteration, our hierarchical search improves the possibility of obtaining the most discriminant feature subset. Another contribution is that we take into account domain knowledge to guide the feature selection. Domain knowledge is often available in the medical field (e.g. organ shape [24], patient characteristics, ...). With respect to predictive factors for patient's outcome, many researches have shown that SUV-based features are of great significance [2–5]. We thus propose to incorporate this knowledge into the hierarchical forward selection, by setting the first feature as a SUV-based features. The prior knowledge constrained feature selection obtains a robust feature subset more efficiently.

The rest of this paper is organized as follows. In Section 2, the proposed hierarchical forward selection method and its prior knowledge version are illustrated in detail. In Section 3, experimental results on two real-world datasets are presented and discussed. The last section, Section 4, concludes this study.

2. Methods

2.1. Hierarchical forward selection (HFS)

In this section we describe our wrapper forward selection algorithm. It starts from an empty set and then iteratively searches forward the binary feature subset space. During an iteration, all feature subsets which perform better than their predecessors in the previous iteration are retained, using a pre-defined criterion J as the evaluation measure. Let J^0 denote an initial lower-bound on J . The search algorithm relies on the following principle, described thereafter: at the k th ($k > 0$) iteration,

- generate the successors of the subsets retained in the previous iteration;
- compute the J values of the successors and select the ones with a better J value than the feature subsets in the previous iteration.

These two phases are performed in turn until J stops increasing. At the last iteration, one or more feature subsets are retained. If more than one feature subsets are retained, we propose an auxiliary criterion to obtain a unique solution. In the following, we will first describe the forward inclusion to generate successors, the selection of outperforming feature subsets, and the stopping criteria along with the auxiliary criterion.

2.1.1. Forward inclusion to generate successors

Let \mathcal{F}_k denote the set of candidate subsets of features obtained at the k th iteration, and \mathcal{G}_{k+1} the set of successors to be evaluated at the following iteration. The successors are generated from

Download English Version:

<https://daneshyari.com/en/article/377569>

Download Persian Version:

<https://daneshyari.com/article/377569>

[Daneshyari.com](https://daneshyari.com)