Computing group cardinality constraint solutions for logistic regression problems

Yong Zhang [a,*], Dongjin Kwon [a,b], Kilian M. Pohl [a,b]

[a] *Department of Psychiatry & Behavioral Sciences, Stanford University, Palo Alto, CA 94304, USA*
[b] *Center for Health Sciences, SRI International, Menlo Park, CA 94025, USA*

ARTICLE INFO

ABSTRACT

We derive an algorithm to directly solve logistic regression based on cardinality constraint, group sparsity and use it to classify intra-subject MRI sequences (e.g. cine MRIs) of healthy from diseased subjects. Group cardinality constraint models are often applied to medical images in order to avoid overfitting of the classifier to the training data. Solutions within these models are generally determined by relaxing the cardinality constraint to a weighted feature selection scheme. However, these solutions relate to the original sparse problem only under specific assumptions, which generally do not hold for medical image applications. In addition, inferring clinical meaning from features weighted by a classifier is an ongoing topic of discussion. Avoiding weighing features, we propose to directly solve the group cardinality constraint logistic regression problem by generalizing the Penalty Decomposition method. To do so, we assume that an intra-subject series of images represents repeated samples of the same disease patterns. We model this assumption by combining series of measurements created by a feature across time into a single group. Our algorithm then derives a solution within that model by decoupling the minimization of the logistic regression function from enforcing the group sparsity constraint. The minimum to the smooth and convex logistic regression problem is determined via gradient descent while we derive a closed form solution for finding a sparse approximation of that minimum. We apply our method to cine MRI of 38 healthy controls and 44 adult patients that received reconstructive surgery of Tetralogy of Fallot (TOF) during infancy. Our method correctly identifies regions impacted by TOF and generally obtains statistically significant higher classification accuracy than alternative solutions to this model, *i.e.*, ones relaxing group cardinality constraints.

1. Introduction

An important topic in medical image analysis is to identify image phenotypes by automatically classifying time series of 3D Magnetic Resonance Images (MRIs). For example, intra-subject MRI sequences are used to analyze cardiac motion (Osman et al., 1999; Sermesant et al., 2003; Chandrashekara et al., 2004; Huang et al., 2005; Besbes et al., 2007; Sundar et al., 2009; Zhang et al., 2010a; Margeta et al., 2012; Wang et al., 2012; Yu et al., 2014), and brain development (Chetelat et al., 2005; Zhang et al., 2010b; Aljabar et al., 2011; Serag et al., 2012; Toews et al., 2012; Bernal-Rusiel et al., 2013; Schellen et al., 2015). However, the automatic classification of medical images is generally challenging. First, the number of features extracted from medical images is usually much larger

than the number of samples. This generally results in overfitting of the method to the data, *i.e.*, much higher classification accuracy during training than on test data (Ryali et al., 2010; Marques et al., 2012; Deshpande et al., 2014). In addition, the image phenotypes identified by automatic classifiers are often difficult to relate to the medical literature (Qu et al., 2003). In this article, we propose an algorithm that addresses both issues by directly solving the so called *logistic regression problem with group sparsity constraints*.

Classifiers based on sparse models reduce the dense image data to a small number of features by counting the number of selected features via the l_0-"norm" and are configured so that the count is below a predefined threshold (Yamashita et al., 2008; Carroll et al., 2009; Rao et al., 2011; Liu et al., 2012; Lv et al., 2015). A generalization of that concept are group sparsity models, which first group image features based on predefined rules and then count the number of non-zero groupings (Ng et al., 2010; Wu et al., 2010; Ryali et al., 2010). To solve the underlying minimization problem, however, these methods relax the feature selection process

* Corresponding author.
  E-mail address: michaelzhang917@gmail.com (Y. Zhang).

from (group) cardinality constraints to weighting feature by, for example, replacing the l_0-"norm" with the l_2-norm (Meier et al., 2008; Friedman et al., 2010; Ryali et al., 2010; Li et al., 2012). The solution of those methods relates to the original sparse problem only under specific assumptions, e.g., the data entry matrix needs to satisfy the restricted isometry property in compressed sensing problem (Candes and Tao, 2005; Candès et al., 2006). However, matrices generally do not satisfy this property, such as those of the appendix of (Lu and Zhang, 2013), and most data matrices of medical image applications, e.g., matrices defined by the regional volume scores of subjects. Thus, with the exception of sparse models applied to compressed sensing, the solution obtained with respect to the relaxed norm generally does not recover the one of the original sparse model defined by the l_0-"norm". In addition, the number of measures selected by the classifier depends now on the training data due to the soft selection scheme. One can select a predefined number by choosing measures whose weight is above a certain threshold. However, in the case of sparse logistic regression the corresponding classifier depends on the measures below the threshold and the relevance of those weights with respect to the disease under study is an ongoing topic of discussion (Haufe et al., 2014; Sabuncu, 2014). Alternatively, the upper bound associated with the sparse constraint is set so that the classifier returns the wanted number of measures for a given training data set (Vounou et al., 2012; Zhang and Shen, 2012; Ma and Huang, 2008; Zhang et al., 2012). The tuning is now data dependent, *i.e.*, each training set is generally associated with a different upper bound so that selected number of scores is constant across training sets. Even comparing the patterns of different subsets of the same data set, *i.e.*, folds, is none trivial as each pattern is the solution to a minimization problem, whose sparsity constraint is unique to a fold. Avoiding soft feature selection and thus these issues, our algorithm solves the original group sparsity constrained, logistic classification problem defined by the l_0-"norm" by extending the Penalty Decomposition (**PD**) method (Lu and Zhang, 2013). By doing so, our method uses a single model to not only classify samples but also directly select patterns (without thresholding or changing upper bounds) that potentially are image phenotypes meaningful to medical community.

To further investigate its potential, we now generalize PD from solving sparse logistic regression problems with group size one to more than one. Specifically, we assume that an intra-subject series of images represents repeated samples of the same disease patterns. In other words, selecting an image feature for disease identification needs to account for the entire series of measurements created by that feature across time. We model this assumption by combining each "feature series" into a single group. The proposed PD algorithm then derives a solution within that model by decoupling the minimization of the logistic regression function from enforcing the group sparsity constraint. Applying Block Coordinate Descent (**BCD**), the minimum to the smooth and convex logistic regression problem is determined via gradient descent while we derive a closed form solution for finding a sparse approximation of that minimum.

We apply our method to cine MRI of 38 healthy adults and 44 adult patients that received reconstructive surgery of Tetralogy of Fallot (**TOF**) during infancy. The data sets fulfill the assumption of the group sparsity model as the residual effects of TOF mostly impact the shape of the right ventricle (Atrey et al., 2010; Bailliard and Anderson, 2009) so that the regions impacted by TOF should not change across the time series captured by a cine MRI. During training, we automatically set all important parameters of our approach by first training a separate regressor for each setting of the parameter space. We then reduce the risk of overfitting by combining those classifiers into a single *ensemble of classifiers* (Rokach, 2010). This ensemble of classifiers correctly favors subregions of the ventricles most likely impacted by TOF. For most experiments, it also produces statistically significant higher accuracy scores than ensemble of classifiers that relax the group cardinality constraint.

We first proposed to generalize PD to group sparsity constraints at MICCAI 2015 (Zhang and Pohl, 2015). This article provides a more in-depth view of this idea. Specifically, we expand PD to guarantee convergence of the sparse approximation to a local minimum of the group-sparsity confined, logistic regression problem, which is the primary contribution of this manuscript. We also modify the experiments by replacing the morphometric encodings of heart regions based on the average of the Jacobian determinants with simple volumetric scores. This simplifies preprocessing as alignment of each cine MRI to a template is unnecessary. It also reduces the size of the parameter search space, which now omits the smoothing parameters associated with the alignment process. Moreover, we not only record a single accuracy score for each implementation but instead generate distributions of scores by modifying the number of training samples. For each training size, we apply the method to 10 different training and testing sets. Finally, we distinguish the ventricular septum from the left ventricle to refine our findings from the previous publication (Zhang and Pohl, 2015) and support those findings with new plots that visualize the selection of regions across the entire heart.

Beyond our MICCAI publication, a possible alternative regression approach for simultaneous classification and pattern extraction is the random forest method (Lempitsky et al., 2009). However, it is unclear how to expand this technology to group-wise selection schemes that enforce temporal consistency in selecting regions, *i.e.*, the same regions are picked across all time points. Due to these difficulties most machine learning approaches applied to cine MRI just focus on disease classification, such as (McLeod et al., 2013; Afshin et al., 2014; Bai et al., 2015). They often improve results by manually selecting regions thought to be impacted by the disease before performing classification (Wald et al., 2009). An exception are (Qian et al., 2011; Ye et al., 2014; Bhatia et al., 2014), which separately perform disease classification and weigh individual regions possibly impacted by disease. The disconnect between the two steps and the weighing of individual regions makes clinical interpretation of the findings more difficult as, in addition to the earlier mentioned issues associated with the interpretation of weights, it increases the risk of false positive findings compared to directly identifying patterns of regions. Our experimental results echo these issues, where logistic regression with relaxed sparsity constraints was generally significantly less accurate than our proposed solution to the original sparsity constraint. We conclude that our proposed approach is the first to solve a single optimization problem for simultaneous disease classification and group-based pattern identification based on segmentation of cine MRIs.

The rest of this paper is organized as follows. Section 2 provides an in-depth description of PD algorithm and its convergence properties. Section 3 summarizes the experiments on the TOF dataset and Section 4 concludes the paper with final remarks.

2. Solving sparse group logistic regression

We first describe the logistic regression model with group cardinality constraint, which accurately assigns subjects to cohorts based on features extracted from intra-subject image sequences. We then generalize the PD approach to find a solution within that model. We end the section deriving convergence properties of the resulting algorithm.

2.1. The model

The input to our model are N subjects, their diagnosis $\{b_1, \ldots, b_N\}$ and features $\{Z^1, \ldots, Z^N\}$ extracted from 3D+t medi-