ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 15, No. 1, 2020, *pp*.016-021

Robust nonlinear multimodal classification of Alzheimer's disease based on GMM

Ziyue Wang[1]
School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China
(Received September 07, 2019, accepted October 26, 2019)

Abstract: Accurate diagnosis of Alzheimer's disease (AD) and its prodromal stage mild cognitive impairment (MCI) is very important for patients and clinicians. There are many useful medical data have been discovered to be remarkable for diagnosis i.e., structural MR imaging (MRI), functional imaging (e.g., FDG-PET and FIB-PET). Multimodal classification model is needed to combine these biomarkers to improve the diagnose performance. Some methods have been proposed such as linear mixed kernel, combined embedding and nonlinear graph fusion. These methods have efficiently employed the multimodal data, but they ignore the influence of noise and outliers. Noise is easily generated in image analysis and measurement. To enhance robustness, mixture distributions were applied in nonlinear regression models. Gaussian mixture model is successfully applied in many domains. In this paper, we generalize nonlinear multimodal classification model based on GMM. The performance on real dataset: 22 AD, 23 MCI and 25 NC (health) is comparable to other methods.

Keywords: Robust nonlinear regression, Outlier, Kernel method, Classification.

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia in elderly people. AD greatly affects the cognitive ability of the elderly [1]. Thus, it is important to diagnose AD as soon as possible from its early stage mild cognitive impairment MCI. In the clinic, many medical images and biological indicators are used for diagnosis. Such as MRI (MR image) [2], functional imaging (FDG-PET, FIB-PET) [3] and quantification of specific proteins measured through CSF [2].

Different biomarkers can contain different feature of AD patient, thus may provide complementary information for diagnosis [4-6]. In [5, 7], linear mixed kernel is proposed independently. Paper [5] learn the kernel weight by grid search while paper [7] take the kernel weight as model parameter and learn it by optimization. Similarities from multiple modalities are combined to generate an embedding, which contain information of multimodal data [6]. In paper [4], similarity matrix for classification is calculated by nonlinear graph fusion. In this paper, kernel method is also used for multimodal data, and the construction of the combined kernel matrix is the same as the mixed kernel of paper [5].

After the construction of the mixed kernel, which contains sample information completely, efficient classification is needed. There are many classification models have been proposed such as logistic regression, k-nearest neighbor, naïve bayes, decision tree, SVM [8-10] and so on. However, most of those classification models do not model the noise directly except support vector machine. Support vector machine model the noises and outliers with the slack variable. In SVM, the input data is mapped into a higher dimensional space to make it separable. SVM can solve two-class classification, and the goal is to maximize the decision bound. This method is totally influenced by the support vectors on the decision bound, if most of those support vectors are polluted by noises, the model will be not proper enough. Therefore, the slack variable is proposed to make the decision bound more robust. Kernel method is improved to deal with the nonlinear case.

Based on the traditional SVM, Least-Squares SVM (LSSVM) [11] is proposed. The LSSVM changes the equality constraint in SVM to the inequality constraint. As a result, the convex quadratic programming is replaced to convex linear problem. In Least-Squares SVM, the slack variables are proportional to the errors.

Mixture models are successfully applied in many domains due to their excellent robustness. In paper [12, 13], mixture of t and skew normal distribution is applied separately to fit the noise term in the linear

[1] Corresponding author. E-mail address: 824831789@qq.com.

regression model. In [14, 15], Gaussian mixture models (GMMs) based classification models are applied in medical research. Paper [16] uses the Gaussian mixture models (GMMs) for multiple limb motion classification using continuous myoelectric signals. Besides, mixture model applied in machine learning [17].

In real world, the data is usually polluted by outliers and heavy-tailed noises, the slack variable in Least-Squares SVM can't be well characterized. In this paper, we develop a nonlinear classification model while the feature of noise is fitted by Gaussian mixture model (GMM). The linear mixed kernel method is employed which contains the multimodal data. In order to get the optimal parameter, EM algorithm and Lagrange multiplier method are applied. The experiment results are comparable to other multimodal-based classification methods.

2. Methodology

2.1 Nonlinear classification model

Given the training set $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, where $\boldsymbol{x}_i \in R^d$ is the input data , $y_i \in \{-1,\ 1\}$ is the label. The objective function of support vector machine is:

$$f(x) = sign[\sum_{i=1}^n \alpha_i k(x_i, x) + b] \tag{1}$$

where α_i is the parameter of Lagrange multiplier method, $k(x_i, x)$ is the kernel function, b is the bias.

Assuming that

$$\begin{cases} \boldsymbol{\omega}^T \phi(\boldsymbol{x}_i) + b = 1 - e_i \\ \boldsymbol{\omega}^T \phi(\boldsymbol{x}_i) + b = -1 + e_i \end{cases}$$

then, we have

$$(\boldsymbol{\omega}^T \phi(\boldsymbol{x}_i) + b) y_i = 1 - e_i, \quad i = 1,2,\cdots,n$$

where $\phi(\cdot)$is the map function, e_i is the error.

The objective function of SVM is:

$$\begin{aligned} &\min_{w,e} \ \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{e}\|_2^2 \\ &s.t. \quad \boldsymbol{G}^T\boldsymbol{\omega} + b\boldsymbol{y} = \boldsymbol{1}_n - \boldsymbol{e} \end{aligned} \tag{2}$$

where $\boldsymbol{G} = (y_1\phi(x_1), y_2\phi(x_2), \dots, y_n\phi(x_n)) \in R^d \times n$, $\boldsymbol{e} = (e_1, e_2, \dots, e_n)$ is the error, γ is a regularized parameter.

2.2 GMM based nonlinear classification model

Gaussian mixture model:

$$p(e) = \sum_{k=1}^K \pi_k N(e|0, \sigma_k^2) \tag{3}$$

where K is the number of independent Gaussian distribution in GMM model. $N(e|0, \sigma_k^2)$ is the Gaussian distribution with zero mean, variance σ_k^2, π_k is the weight coefficient that satisfied: $\sum_{k=1}^K \pi_k = 1,\ \pi_k \geq 0$.

Theoretically, we need to define the form of the map function $\phi(x)$ in advance. However, this will increase the number of coefficient and computation complexity, in the same time, choosing a mapping function is complicated. Similar to LS-SVM, we will use the Lagrange multiplier method in the optimize step. So the map function always appears as $\phi(x)^T\phi(x)$. Therefore, we can introduce kernel function $k(x, y) = \phi(x)^T\phi(y)$. In this paper, RBF kernel is employed:

$$k(x, y) = exp(-\frac{\|x-y\|^2}{2\sigma^2}) \tag{4}$$

The optimal values of parameter can be obtained by maximum likelihood estimation, the likelihood function of e can be expressed as:

$$p(e|\Theta) = \prod_{i=1}^n p(e_i|\Theta) = \prod_{i=1}^n \sum_{k=1}^K \pi_k N(0, \sigma_k^2) \tag{5}$$

where Θ is the parameter set. Then the log-likelihood function is calculated as:

$$L(e|\Theta) = \sum_{i=1}^n log\, p\,(e_i|\Theta) = \sum_{i=1}^n (log \sum_{k=1}^K \pi_k N(0, \sigma_k^2)) \tag{6}$$

Due to the complex expression of log-likelihood function, it is difficult to calculate directly. The EM algorithm is an efficient algorithm to solve such problems.

In order to simplify the solution process, we introduce $\boldsymbol{Z} = (z_1, z_2, \dots, z_n)^T$, where $z_i = (z_{i1}, z_{i2}, \dots, z_{iK})$ is an indicator vector, if e_i comes from the jth component, then $z_{ij} = 1$ the other elements of z_i are 0. So $\sum_{k=1}^K z_{ik} = 1$, $\sum_{i=1}^n \sum_{k=1}^K z_{ik} = 1$.

z_i obeys multi-point distribution: