Journal of the Taiwan Institute of Chemical Engineers 81 (2017) 31–39

Contents lists available at ScienceDirect

Journal of the Taiwan Institute of Chemical Engineers

journal homepage: www.elsevier.com/locate/jtice

Detection of minimum biomarker features via bi-level optimization framework by nested hybrid differential evolution

Kai-Cheng Hsu[a], Feng-Sheng Wang[b,*]

[a] Department of Neurology, National Taiwan University Hospital, Yunlin Branch, Yunlin 64041, Taiwan
[b] Department of Chemical Engineering, National Chung Cheng University, Chiayi 62102, Taiwan

ARTICLE INFO

Article history:
Received 19 July 2017
Revised 11 October 2017
Accepted 13 October 2017
Available online 8 November 2017

Keywords:
Precision medicine
Machine learning
Biomarker detection
Evolutionary optimization
Bilevel optimization

ABSTRACT

One of the important tasks in precision medicine is to identify biomarkers and build classification models for clinical diagnose and treatment response. Support vector machine using full features is a common approach for classifying diseases in healthcare systems. However, little literature reported to use it toward determining minimum features of biomarkers. This study introduced a bilevel mixed-integer optimization framework to detect minimum biomarker features for support vector machine. We proposed the two-population nested hybrid differential evolution (NHDE) to solve the optimization problem for selecting the desired biomarkers. In case studies, the accuracies of classification by SVM using full biomarkers were nearly identical to that of 2 biomarkers selected by the minimizing feature approach. Furthermore, the approach could determine that the dopamine packed in vesicle in the presynaptic dopamine overactivity case and S-adenosyl-L-homocysteine in deficient case were the dominant biomarkers, respectively. The two-population NHDE algorithm was more efficient to achieve minimum biomarkers compared with one-population NHDE and traditional genetic algorithm.

1. Introduction

A bi-level optimization problem (BLOP) is a mathematical programming problem that involves two levels of optimization tasks [1–3]. BLOPs are different from the common optimization problems, as they contain a nested optimization task within the constraints of another optimization problem. The outer optimization problem is referred as the upper level task and the inner optimization problem is referred as the lower level task. The investigations in BLOPs are strongly motivated by real-world applications found in economics, engineering, transportation, networks, planning, and computational biology [4,5]. A BLOP is a special type of multi-objective optimization (MOO) problem, and the objectives between the upper and lower level are hierarchical relationships so that conventional MOO methods cannot be directly applied to solve BLOPs. Numerous conventional algorithms have been proposed to solve BLOPs; these algorithms can be classified into three categories: vertex enumeration, Kuhn–Tucker and evolutionary algorithms [2,6]. Vertex enumeration algorithms cannot solve large-scale problems because they use brute-force computation to enumerate every integer variable. Kuhn–Tucker algorithms have been employed to reduce a BLOP to a single-level optimization problem by using strong duality theory. However, the computation time when such an approach is used can increase exponentially when the number of decision variables is increased.

A few studies have considered solving BLOPs through evolutionary optimization, and most of the methods proposed are nested in nature, as discussed on the website of Evolutionary Bi-level Optimization [6]. One of the earliest evolutionary algorithms for solving BLOPs was proposed by Mathieu et al. [3], who used a genetic algorithm at the outer level and linear programming at the inner level. Yin [7] solved the outer-level sub-problem by using a genetic algorithm and the inner-level sub-problem by using the reduced gradient method. Differential evolution (DE) at both levels and nested DE with ant colony optimization have also been applied to solve BLOPs [1,8,9]. However, such algorithms require lot of computations to determine an optimal solution for large-scale BLOPs, such as the rational strain design problem of genome-scale metabolic networks. Wang and Wu [10] have introduced a nested hybrid differential evolution (NHDE) to solve a genome-scale growth-coupled production strain design problem to overcome such a drawback. One population of individuals was employed in the NHDE algorithm to determine minimum number of knocked out genes. However, the premature minimum number

* Corresponding author.
E-mail address: chmfsw@ccu.edu.tw (F.-S. Wang).

https://doi.org/10.1016/j.jtice.2017.10.015
1876-1070/

Notations

b	the bias
C	the penalty parameter
$d(\mathbf{x})$	the linear discriminant function
$F(\boldsymbol{\alpha}, \mathbf{z})$	the outer/leader objective
$f(\boldsymbol{\alpha}, \mathbf{z})$	the inner/follower objective
$\mathbf{G}(\boldsymbol{\alpha}, \mathbf{z})$	the vectors of inequality constraints in the outer optimization problems
$\mathbf{g}(\boldsymbol{\alpha}, \mathbf{z})$	the vectors of inequality constraints in the inner optimization problems
$\mathbf{H}(\boldsymbol{\alpha}, \mathbf{z})$	the vectors of equality constraints in the outer optimization problems
$\mathbf{h}(\boldsymbol{\alpha}, \mathbf{z})$	the vectors of equality constraints in the inner optimization problems
$K(\mathbf{x}_i, \mathbf{x}_j)$	the kernel function
T_i	the prediction indicator
$\mathbf{w}$	the weight vector of the discriminant function
$\mathbf{x}_i$	the ith features or input
y_i	the negative/positive pattern for the ith training data
$\mathbf{z}$	an n-dimensional vector of integer variables
z_k	the binary variables for the kth feature
α_i	the ith Lagrange multiplier
ξ_i	the slack variable for the ith training data
Ω^{TN}	the set of the training data
Ω^{TS}	the set of the testing data

could be achieved by the one population approach. This study proposes two populations of individuals in the NHDE algorithm to surmount the weakness.

In clinical practice, physicians make diagnosis based on the symptoms and signs of patients. However, different disorders often manifest similarly and overlap of symptoms and signs cause diagnostic difficulty. Biomarkers using molecular biology techniques [11] may increase the accuracy of diagnosis and allow disease classifications effectively targeted for precision medicine [12]. The most biomarkers are generally designed by a support vector machine (SVM) using full features. However, such a SVM should use all features of a patient so that the diagnostic cost is expensive. Moreover, it is difficult to understand which feature or metabolite is dominant in the system because some features are correlated and dependent. This study therefore introduced a minimizing biomarker framework toward reducing diagnostic costs with similar accuracies and achieving the dominant features.

2. Bilevel optimization for biomarker detection

2.1. Support vector machine

Support vector machine (SVM) is a useful technique for data classification introduced by Vapnik [13]. The SVM classifier is widely used in bioinformatics and computational biology and other disciplines due to its high accuracy, ability to deal with high-dimensional data such as gene expression, and flexibility in modeling diverse sources of data [14–21]. SVM is a popular supervised learning method that is a means to divide data into categories based on common characteristics through a single-level optimization. The data for training a SVM consist of a set of classes or objects with its corresponding labels. Each objective contained many features or attributes obtained from observations. For instance, we have L training data, where each input $\mathbf{x}_i$ has n attributes and is in one of two classes $y_i = -1$ (negative pattern) or $y_i = +1$ (positive pattern), i.e. our training data is of the form

$$\{(\mathbf{x}_i, y_i), i = 1, ..., L\}, \text{ where } y_i \in \{-1, +1\}, \mathbf{x} \in \Re^n \tag{1}$$

A linear classifier is based on a linear discriminant function of the form

$$d(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{2}$$

where the vector $\mathbf{w}$ is the weight vector and the scalar b is the bias. The linear discriminant function is a hyperplane, i.e. a linear decision boundary, to linearly separate the dataset into positive or negative patterns. For a given hyperplane, we can determine the maximum margin classifier that maximizes the geometric margin, which is equivalent to minimizing the soft-margin SVM problem of the form

$$\begin{cases} \min\limits_{\mathbf{w},b,\xi_i} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum\limits_{i=1}^{L}\xi_i \\ \text{subject to} \\ y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i, i = 1, ..., L \\ \xi_i \geq 0 \end{cases} \tag{3}$$

where the penalty parameter $C > 0$ assigns the relative importance of maximizing the margin and minimizing the amount of slack variables ξ_i penalty. Once the optimal values of $\mathbf{w}$, b and ξ_i are found, the patterns in the training set can be classified by the discriminant function $d(\mathbf{x})$ either into the -1 or $+1$ class. Using the method of Lagrange multipliers, the problem (3) can be formulated as the dual problem, which is expressed in terms of Lagrange multipliers α_i

$$\begin{cases} \min\limits_{\alpha} \left[\frac{1}{2}\alpha^T\mathbf{H}\alpha - \mathbf{e}^T\alpha\right] \\ \text{subject to} \\ \sum\limits_{i=1}^{L}\alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C, \ i = 1, ..., L \end{cases} \tag{4}$$

where $\mathbf{H} = [H_{ij}] = [y_i \ y_j \ K(\mathbf{x}_i,\mathbf{x}_j)]$ and $\mathbf{e}^T = [1,...,1]$. The linear kernel $K(\mathbf{x}_i,\mathbf{x}_j)$ is the inner product of $\mathbf{x}_i$ and $\mathbf{x}_j$. The dual formulation is a convex quadratic optimization problem, and we can use a QP solver to find α_i, and leads to an expansion of the weight vector in terms of the input features

$$\mathbf{w} = \sum_{i=1}^{L}\alpha_i y_i \mathbf{x}_i \tag{5}$$

2.2. Minimization of biomarker features

SVM has applied for biomarker identification problems using all features of each class [14,16–23]. However, SVM using all features is impractical and expensive diagnostic examination. Furthermore, it is unable to determine the dominant feature because some features are correlated and dependent in the system. Two types of feature selection methods, filter method and wrapper method [24–26], are applied to limit the number of input feature in classifier in order to have high performance accuracy. Both methods do not consider the minimization of features as an objective in SVM so that minimum biomarkers may not be achieved. This study proposes a bilevel mixed-integer optimization problem (BLMIOP) to determine the minimum features to achieve the optimal classification. This BLMIOP is referred to as the minimizing biomarker