Information Processing Letters 115 (2015) 833–837

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

On the approximability of the Largest Sphere Rule Ensemble Classification problem ☆

Vahan Mkrtchyan, K. Subramani *,[1]

Lane Department of Computer Science and Electrical Engineering, West Virginia University, United States

ARTICLE INFO

Article history:
Received 11 June 2014
Received in revised form 13 June 2015
Accepted 16 June 2015
Available online 22 June 2015
Communicated by B. Doerr

Keywords:
Approximation algorithms
Weighted Rule Ensemble Classification
Reversal problem
Feasibility
NP-completeness
Approximation
Inapproximability

ABSTRACT

In this paper, we discuss the approximability of the Largest Sphere Rule Ensemble Classification problem (**LSRES**). This problem is closely related to the Weighted Rule Ensemble Classification Reversal problem (**WRECR**) [2].

1. Introduction

In this paper, we discuss the approximability of the Largest Sphere Rule Ensemble Classification problem (**LSRES**). This problem is closely related to the Weighted Rule Ensemble Classification Reversal problem (**WRECR**) [2].

WRECR is a well-known problem, widely referenced in the literature. This problem has applications in domains such as biology, medicine and computer security [2,3,5,6]. In this paper, we introduce a related problem called **LSRES**, and motivate it through an application in computer security (see Section 4).

The main contributions of this paper are as follows:

(i) The design of a $\frac{C}{n}$-approximation algorithm for **LSRES**, with running time $O(n^{C+1})$ for each fixed $C > 0$.
(ii) The proof of absence of a polynomial-time $\frac{1}{n^{1-\epsilon}}$-approximation algorithm for **LSRES**, unless **P**=**NP**. This result holds for each $\epsilon \in (0, 1)$.
(iii) The proof of **NP-completeness** of the feasibility version of **WRECR**.

The rest of the paper is organized as follows: In Section 2, we discuss the preliminaries and notations used throughout the paper. Section 3 presents the formulations of the main problems discussed in the paper. Section 4 discusses a problem that motivates the study of **LSRES** and presents related work in the literature. Section 5 discusses the main results obtained in this paper. In Section 6, we summarize our contributions and outline directions for future research.

2. Preliminaries

In this paper, we use the standard convention that 0 corresponds to **false**, and 1 corresponds to **true**. Let

☆ This research has been supported in part by the Air Force Office of Scientific Research through Award FA9550-12-1-0199.

* Corresponding author.
E-mail addresses: vahan.mkrtchyan@mail.wvu.edu (V. Mkrtchyan), ksmani@csee.wvu.edu (K. Subramani).

[1] The author is supported by the National Science Foundation through Award CCF-1305054.

http://dx.doi.org/10.1016/j.ipl.2015.06.012
0020-0190/© 2015 Elsevier B.V. All rights reserved.

$B = \{0, 1\}$. The operations defined on B are algebraic or Boolean depending on the context.

A Boolean function f is a mapping $f : B^n \to B$, where $B^n = B \times B \times \ldots \times B$.

Definition 1. Let $f(x_1, \ldots, x_n)$ be a Boolean function. Its dual Boolean function [7] is defined as:

$$f^*(x_1, \ldots, x_n) = \bar{f}(\bar{x}_1, \ldots, \bar{x}_n).$$

From the definition, it follows that for any Boolean function f, we have,

$$(f^*)^* = f.$$

For example, if

$$f_1(x) = x,\ f_2(x) = \bar{x},\ \text{and}\ f_3(x, y) = x \wedge y,$$

then

$$f_1^*(x) = x,\ f_2^*(x) = \bar{x},\ \text{and}\ f_3^*(x, y) = x \vee y.$$

The first two equalities show that the identity and complement functions are self-dual (i.e., the function coincides with its dual). The last equality can be generalized to show that the dual of a conjunction is a disjunction and vice versa.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an n-dimensional vector over $\{0, 1\}$. A decision rule r is a conjunction of literals of the variables $x_1, \ldots, x_n$. Assume, we are given a set of decision rules $R = \{r_1, \ldots, r_t\}$, and a weight-function $w : R \to \mathbb{Z}$. Consider the expression $\Delta(R, w, \mathbf{x})$ defined as follows:

$$\Delta(R, w, \mathbf{x}) = \sum_{i=1}^{t} w(r_i) \cdot r_i(\mathbf{x}).$$

In the expression above we interpret the values of rules algebraically.

Let $F(\mathbf{x})$ be a function defined as:

$$F(\mathbf{x}) = sgn(\Delta(R, w, \mathbf{x})), \tag{1}$$

where

$$sgn(z) = \begin{cases} 1, & \text{if } z > 0, \\ -1, & \text{if } z < 0, \\ 0, & \text{otherwise.} \end{cases}$$

For a Boolean vector $\mathbf{x} = (x_1, \ldots, x_n)$ and a set $S \subseteq \{1, \ldots, n\}$, let $\mathbf{x}^S$ be the vector obtained from $\mathbf{x}$ by complementing the elements with indices from S.

The Weighted Rule Ensemble Classification Reversal problem (**WRECR**) is defined as follows:

WRECR: Given a set of rules $R = \{r_1, \ldots, r_t\}$, a weight function $w : R \to Z$, a Boolean vector $\mathbf{x} = (x_1, \ldots, x_n)$ and a positive integer $k \le n$, is there a set $S \subseteq \{1, \ldots, n\}$ with $|S| \le k$, such that $F(\mathbf{x}^S) \ne F(\mathbf{x})$?

If we partition the set of n-dimensional Boolean vectors into 3 groups based on the values of F (see Fig. 1), then the **WRECR** problem can be interpreted as the problem of checking whether there is a set $S \subseteq \{1, \ldots, n\}$ with $|S| \le k$, such that the vector $\mathbf{x}^S$ belongs to a different group (see Fig. 1).

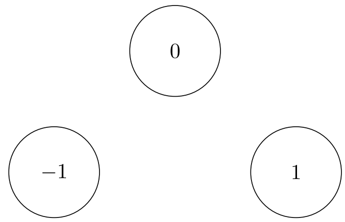

Fig. 1. Partitioning n-dimensional Boolean vectors according to the values of F.

In this paper, we derive some results that are concerned with the approximation properties of a problem that is related to **WRECR**.

3. Formulation of the main problems

We first consider the feasibility version of **WRECR**:

WRECRF: Given a set of rules $R = \{r_1, \ldots, r_t\}$, a weight function $w : R \to Z$, and a Boolean vector $\mathbf{x} = (x_1, \ldots, x_n)$, is there a set $S \subseteq \{1, \ldots, n\}$, such that $F(\mathbf{x}^S) \ne F(\mathbf{x})$?

Observe that **WRECR** is the decision problem corresponding to the problem of constructing a smallest cardinality feasible set S if one exists.

Now, we consider the **LSRES** problem, which is defined as follows:

LSRES: Given a set of rules $R = \{r_1, \ldots, r_t\}$, a weight function $w : R \to Z$ and a Boolean vector $\mathbf{x} = (x_1, \ldots, x_n)$, the goal is to find the maximum $d \ge 0$, such that for every set $S \subseteq \{1, \ldots, n\}$ with $|S| \le d$, $F(\mathbf{x}^S) = F(\mathbf{x})$ (see (1)).

Observation 1. LSRES *is always feasible. In other words, in this problem the parameter d is well-defined. One can always take $d = 0$, in order to get a feasible solution.*

Lemma 1. LSRES *is as hard as* **WRECR** *(***WRECR** $\prec$ **LSRES***).*

Proof. We Turing-reduce **WRECR** to **LSRES**. Let $P = \langle R, w, \mathbf{x}, k\rangle$ be an instance of **WRECR**. We construct the instance $Q = \langle R, w, \mathbf{x}\rangle$ of **LSRES**. Assume that there is an algorithm which determines the parameter d in Q. Then it suffices to check whether $(d + 1) \le k$ in order to solve P. □

Finally, consider the decision version of **LSRES**:

LSRESD: Given a set of rules $R = \{r_1, \ldots, r_t\}$, a weight function $w : R \to Z$, a Boolean vector $\mathbf{x} = (x_1, \ldots, x_n)$ and a number l, the goal is to check whether for every set $S \subseteq \{1, \ldots, n\}$ with $|S| \le l$, $F(\mathbf{x}^S) = F(\mathbf{x})$.

Lemma 2. LSRESD *is in* **coNP***.*

Proof. Assume that we are given a set of rules $R = \{r_1, \ldots, r_t\}$, a weight function $w : R \to Z$, a Boolean vector $\mathbf{x} = (x_1, \ldots, x_n)$, a number l and a set $S \subseteq \{1, \ldots, n\}$ with $|S| \le l$ and $F(\mathbf{x}^S) \ne F(\mathbf{x})$. We can verify that S satisfies $|S| \le l$ and $F(\mathbf{x}^S) \ne F(\mathbf{x})$ in polynomial time. Thus, negative instances of **LSRESD** have a certificate of polynomial length, which can be checked in polynomial time. □