Discrete Applied Mathematics 179 (2014) 13–27

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

Influence and interaction indexes for pseudo-Boolean functions: A unified least squares approach

Jean-Luc Marichal [a,*], Pierre Mathonet [b]

[a] Mathematics Research Unit, FSTC, University of Luxembourg, 6, rue Coudenhove-Kalergi, L-1359 Luxembourg, Luxembourg

[b] University of Liège, Department of Mathematics, Grande Traverse, 12 - B37, B-4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 19 January 2012
Received in revised form 10 April 2014
Accepted 18 May 2014
Available online 11 June 2014

Keywords:
Cooperative game
Pseudo-Boolean function
Power index
Influence index
Interaction index
Least squares approximation

ABSTRACT

The Banzhaf power and interaction indexes for a pseudo-Boolean function (or a cooperative game) appear naturally as leading coefficients in the standard least squares approximation of the function by a pseudo-Boolean function of a specified degree. We first observe that this property still holds if we consider approximations by pseudo-Boolean functions depending only on specified variables. We then show that the Banzhaf influence index can also be obtained from the latter approximation problem. Considering certain weighted versions of this approximation problem, we introduce a class of weighted Banzhaf influence indexes, analyze their most important properties, and point out similarities between the weighted Banzhaf influence index and the corresponding weighted Banzhaf interaction index. We also discuss the issue of reconstructing a pseudo-Boolean function from prescribed influences and point out very different behaviors in the weighted and non-weighted cases.

1. Introduction

Let $f\colon \{0,1\}^n \to \mathbb{R}$ be an n-variable pseudo-Boolean function and let S be a subset of its variables. Define the *influence of* S *over* f as the expected value, denoted $I_f(S)$, of the highest variation of f when assigning values independently and uniformly at random to the variables not in S (see [12] for a normalized version of this definition). That is,

$$I_f(S) = \frac{1}{2^{n-|S|}} \sum_{T\subseteq N\setminus S} \Big(\max_{R\subseteq S} f(T\cup R) - \min_{R\subseteq S} f(T\cup R)\Big),$$

where $N = \{1,\ldots,n\}$.[1] This notion was first introduced for Boolean functions $f\colon\{0,1\}^n \to \{0,1\}$ by Ben-Or and Linial [2] (see also [10]). There the influence $I_f(S)$ was (equivalently) defined as the probability that, assigning values independently and uniformly at random to the variables not in S, the value of f remains undetermined. Since its introduction, this concept has found many applications in discrete mathematics, cooperative game theory, theoretical computer science, and social choice theory (see, e.g., the survey article [11]).

When the function f is nondecreasing in each variable, the formula above reduces to

$$I_f(S) = \frac{1}{2^{n-|S|}} \sum_{T\subseteq N\setminus S} \big(f(T\cup S) - f(T)\big). \tag{1}$$

* Corresponding author. Tel.: +352 4666446662.
E-mail addresses: jean-luc.marichal@uni.lu (J.-L. Marichal), p.mathonet@ulg.ac.be (P. Mathonet).

[1] Throughout we identify Boolean vectors $\mathbf{x} \in \{0,1\}^n$ and subsets $T \subseteq N$ by setting $x_i = 1$ if and only if $i \in T$. We thus use the same symbol to denote both a pseudo-Boolean function $f\colon\{0,1\}^n \to \mathbb{R}$ and the corresponding set function $f\colon 2^N \to \mathbb{R}$ interchangeably.

http://dx.doi.org/10.1016/j.dam.2014.05.039
0166-218X/

The latter expression has an interesting interpretation even if f is not nondecreasing. In cooperative game theory for instance, where $f(T)$ represents the worth of coalition T in the game f, this expression is precisely the average value of the marginal contributions $f(T \cup S) - f(T)$ of coalition S to outer coalitions $T \subseteq N \setminus S$. Thus, it measures an overall influence (which can be positive or negative) of coalition S in the game f. In particular, when $S = \{i\}$ is a singleton it reduces to the Banzhaf power index

$$I_f(\{i\}) = \frac{1}{2^{n-1}} \sum_{T \subseteq N \setminus \{i\}} \big(f(T \cup \{i\}) - f(T)\big).$$

Thus, the expression in (1) can be seen as a variant of the original concept of influence that simply extends the Banzhaf power index to coalitions. We call it the *Banzhaf influence index* and denote it by $\Phi_{\mathrm{B}}(f, S)$. Actually, this index was introduced, axiomatized, and even generalized to weighted versions in [13].

The *Banzhaf interaction index* [17], another index which extends the Banzhaf power index to coalitions, is defined for a pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ and a subset $S \subseteq N$ by

$$I_{\mathrm{B}}(f, S) = \frac{1}{2^{n-|S|}} \sum_{T \subseteq N \setminus S} (\Delta_S f)(T), \tag{2}$$

where $\Delta_S f$ denotes the S-difference (or discrete S-derivative) of f.[2] When $|S| \geqslant 2$, this index measures an overall degree of interaction among the variables of f that are in S. When f is a game, it measures an overall degree of interaction among the players of coalition S in the game f (see, e.g., [5–7]).

It is known that the Banzhaf power and interaction indexes can be obtained from the solution of a standard least squares approximation problem for pseudo-Boolean functions (see [6,8]). Weighted versions of this approximation problem recently enabled us to define a class of weighted Banzhaf interaction indexes having several nice properties (see [14]). However, we observe that there is no such least squares construction for the Banzhaf influence index in the literature.

In this paper we fill this gap in the following way. In Section 2 we first show that the Banzhaf interaction index can be obtained from a different, more natural (but still elementary) least squares approximation problem. Specifically, $I_{\mathrm{B}}(f, S)$ appears as the leading coefficient in the multilinear representation of the best approximation f_S of f by a pseudo-Boolean function that depends only on the variables in S. We then prove that the Banzhaf influence index $\Phi_{\mathrm{B}}(f, S)$ can be obtained from the same approximation problem simply by considering the difference $f_S(S) - f_S(\varnothing)$. In Section 3 we introduce a class of weighted Banzhaf influence indexes from the solution of a weighted version of this approximation problem. We show that these indexes define a subclass of the family of *generalized values*, give their most important properties, and point out similarities between the weighted Banzhaf influence index and the corresponding weighted Banzhaf interaction index. In Section 4 we discuss the issue of representing pseudo-Boolean functions in terms of Banzhaf influence indexes. More precisely, we show that in the generic weighted case any pseudo-Boolean function can be reconstructed, up to an additive constant, from prescribed influences. By contrast, in the non-weighted case only half of the information contained in the pseudo-Boolean function can be reconstructed. This important observation fully motivates the investigation of the weighted case, which therefore is not a straightforward extension of the non-weighted case. Finally, in Section 5 we present an application of the weighted Banzhaf influence index in system reliability theory and give a couple of concluding remarks.

2. Interactions, influences, and least squares approximations

In this section we recall how the Banzhaf interaction index can be obtained from the solution of a standard least squares approximation problem and we show how a variant of this approximation problem can be used to define both the Banzhaf interaction and influence indexes.

It is well known (see, e.g., [9]) that any pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ can be uniquely represented by a multilinear polynomial function

$$f = \sum_{T \subseteq N} a(T)\, u_T,$$

where $u_T(\mathbf{x}) = \prod_{i \in T} x_i$ is the *unanimity game* (or *unanimity function*) for $T \subseteq N$ (with the convention $u_\varnothing = 1$) and the set function $a\colon 2^N \to \mathbb{R}$, called the *Möbius transform* of f, is defined through the conversion formulas (Möbius inversion formulas)

$$a(S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} f(T) \quad \text{and} \quad f(S) = \sum_{T \subseteq S} a(T). \tag{3}$$

[2] The differences of f are defined as $\Delta_\varnothing f = f$, $\Delta_{\{i\}} f(\mathbf{x}) = f(\mathbf{x} \mid x_i = 1) - f(\mathbf{x} \mid x_i = 0)$, and $\Delta_S f = \Delta_{\{i\}} \Delta_{S \setminus \{i\}} f$ for $i \in S$.