THEORY OF COMPUTING, Volume 6 (2010), pp. 81–84
www.theoryofcomputing.org

NOTE

Decision Trees and Influence: an Inductive Proof of the OSSS Inequality

Homin K. Lee

Received: October 23, 2009; published: August 17, 2010.

Abstract: We give a simple proof of the OSSS inequality (O'Donnell, Saks, Schramm, Servedio, FOCS 2005). The inequality states that for any decision tree T calculating a Boolean function $f : \{0,1\}^n \to \{-1,1\}$, we have $\mathrm{Var}[f] \le \sum_i \delta_i(T)\,\mathrm{Inf}_i(f)$, where $\delta_i(T)$ is the probability that the input variable x_i is read by T and $\mathrm{Inf}_i(f)$ is the influence of the ith variable on f.

ACM Classification: G.2, G.3, F.1.3

AMS Classification: 05A20, 60C05, 68R05, 68Q87

Key words and phrases: computational complexity, decision trees, influence

1 Introduction

Let T be a decision tree computing a function $f : \{0,1\}^n \to \{-1,1\}$. We write $\delta_i(T)$ for the probability that the variable x_i is queried by the decision tree on a uniform random input, and we write:

$$\Delta(T) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \delta_i(T) = \mathop{\mathbb{E}}_{x\in\{0,1\}^n}[\#\text{ coordinates } T \text{ queries on } x]\,.$$

$\Delta(T)$ can also be thought of as the average depth of the decision tree, or as a refinement of the notion of the size of the decision tree, since $\Delta(T) \le \log(\mathrm{size}(T))$ [4].

The influence of a variable x_i on a Boolean function f is defined to be

$$\mathrm{Inf}_i(f) = \Pr_{x\in\{0,1\}^n}[f(x) \ne f(x^{(i)})]\,,$$

DOI: 10.4086/toc.2010.v006a004

where $x^{(i)}$ denotes x with its i-th bit flipped [1]. Recall that the variance of a function f is $\mathrm{Var}[f] = \mathbb{E}[(f-\mathbb{E}[f])^2]$, and that the covariance of two functions f and g is $\mathrm{Cov}[f,g] = \mathbb{E}[(f-\mathbb{E} f)(g-\mathbb{E} g)]$.

O'Donnell et al. [3] proved the following inequality:

Theorem 1.1. *Let $f : \{0,1\}^n \to \{-1,1\}$ be a Boolean function, and let T be a decision tree computing f. Then*

$$\mathrm{Var}[f] \le \sum_{i=1}^{n} \delta_i(T)\,\mathrm{Inf}_i(f)\,.$$

This inequality can be viewed as a refinement of the Efron-Stein Inequality [2, 5] for the discrete cube (i. e., $\mathrm{Var}[f] \le \sum_{i=1}^{n} \mathrm{Inf}_i(f)$) that takes into account the complexity of the function's representation.

O'Donnell et al. [3] also proved the following generalization of Theorem 1.1, which is what we reprove in the next section.

Theorem 1.2. *Let $f,g : \{0,1\}^n \to \{-1,1\}$ be Boolean functions, and let T be a decision tree computing f. Then*

$$|\mathrm{Cov}[f,g]| \le \sum_{i=1}^{n} \delta_i(T)\,\mathrm{Inf}_i(g)\,.$$

2 Inductive Proof

The original proofs of both Theorems 1.1 and 1.2 relied on some delicate probabilistic reasoning about the independence of certain hybrid inputs to the decision tree. We will prove Theorem 1.2 using induction. To do so, we will consider the function's behavior under the two cases when the root variable takes the value 0 and the value 1. First we will review a fact from probability theory.

For a function $f : \{0,1\}^n \to \mathbb{R}$, let

$$c_i(x_1,\dots,x_n) = \mathbb{E}[f|(x_1,\dots,x_i)] - \mathbb{E}[f|(x_1,\dots,x_{i-1})]$$

for $1 \le i \le n$ so that $\mathbb{E}[f] + \sum_i c_i = f$. The sequence $\{c_i\}$ is a martingale difference sequence. Let g be another real-valued function, and let $\{d_i\}$ be its martingale difference sequence. Then $\mathrm{Cov}[f,g] = \sum_{i=1}^{n} \mathbb{E}[c_i d_i]$. We'll prove this fact for the sake of completeness.

Fact 2.1. Let $f,g : \{0,1\}^n \to \mathbb{R}$ be real-valued functions, with martingale difference sequences

$$c_i = \mathbb{E}[f|(x_1,\dots,x_i)] - \mathbb{E}[f|(x_1,\dots,x_{i-1})] \quad \text{and} \quad d_i = \mathbb{E}[g|(x_1,\dots,x_i)] - \mathbb{E}[g|(x_1,\dots,x_{i-1})]$$

for $1 \le i \le n$. Then $\mathrm{Cov}[f,g] = \sum_{i=1}^{n} \mathbb{E}[c_i d_i]$.

Proof. For $j < k$, we have

$$\mathbb{E}[c_j d_k] = \mathbb{E}\,\mathbb{E}[c_j d_k|(x_1,\dots,x_{k-1})] = \mathbb{E}[c_j\,\mathbb{E}[d_k|(x_1,\dots,x_{k-1})]] = \mathbb{E}[c_j \cdot 0] = 0\,.$$

Therefore $\mathrm{Cov}[f,g] = \mathbb{E}[(f-\mathbb{E} f)(g-\mathbb{E} g)] = \mathbb{E}[\sum c_j \sum d_k] = \sum_{i=1}^{n} \mathbb{E}[c_i d_i]$. □

We now relate the last martingale difference sequence with the influence of the variable x_n.

Lemma 2.2. *Let $f,g : \{0,1\}^n \to \{-1,1\}$ be Boolean functions, and let*

$$c_n = f - \mathbb{E}[f|(x_1,\dots,x_{n-1})] \quad \textit{and} \quad d_n = g - \mathbb{E}[g|(x_1,\dots,x_{n-1})]\,.$$

Then $\mathbb{E}[c_n d_n] \le \mathrm{Inf}_n(f)$ and $\mathbb{E}[c_n d_n] \le \mathrm{Inf}_n(g)$.

Proof. Let $f_0(x_1,\dots,x_n)$ denote $f(x_1,\dots,x_{n-1},0)$, and let f_1, g_0, and g_1 be defined similarly. Then we have that

$$c_n = f - (f_0+f_1)/2 \quad \text{and} \quad d_n = g - (g_0+g_1)/2\,.$$

We can rewrite $\mathbb{E}[c_n d_n]$ as

$$\mathbb{E}\left[\left(f - \frac{f_0+f_1}{2}\right)\left(g - \frac{g_0+g_1}{2}\right)\right].$$

If both $f_0 \ne f_1$ and $g_0 \ne g_1$, the quantity inside the expectation is $f \cdot g \in \{-1,1\}$. Otherwise, the quantity is 0.

The influence of x_n on f is just $\mathrm{Inf}_n(f) = \Pr[f_0(x) \ne f_1(x)]$, and thus $\mathrm{Inf}_n(f)$ is an upper bound on $\mathbb{E}[c_n d_n]$ (as is $\mathrm{Inf}_n(g)$). Note that this upper bound is an equality when we consider the special case of $f = g$, and we have that $\mathbb{E}[c_n^2] = \mathrm{Inf}_n(f)$. □

We are now ready to prove Theorem 1.2.

Proof. We'll prove the statement by induction on the number of variables. For the base case of one variable, recall that both $\delta_i(T)$ and $\mathrm{Inf}_i(g)$ are always non-negative, and that the covariance of two functions with range $\{-1,1\}$ is a value in $[-1,1]$. A Boolean function on only one variable is either constant, the single variable x_1, or its negation. If either f or g is constant, then $\mathrm{Cov}[f,g] = 0$, and the inequality holds. If neither f nor g are constant, then $\mathrm{Inf}_1(g)$ and $\delta_1(T)$ must be 1 and the inequality holds.

Now we'll consider f and g on n variables. We can assume that f and g are non-constant, or the inequality trivially holds as before. Thus, T must query at least one variable, and we will assume without loss of generality that the root of T queries x_n. Let T_0 be the left subtree and let T_1 be the right subtree. Then for $i \ne n$, we have $\delta_i(T) = (1/2)\delta_i(T_0) + (1/2)\delta_i(T_1)$. As in the proof of Lemma 2.2, let $f_0(x_1,\dots,x_n)$ denote $f(x_1,\dots,x_{n-1},0)$, and let f_1, g_0, and g_1 be defined similarly. For $i \ne n$, we get the following expression: $\mathrm{Inf}_i(g) = (1/2)\,\mathrm{Inf}_i(g_0) + (1/2)\,\mathrm{Inf}_i(g_1)$.

By Fact 2.1, we can write $\mathrm{Cov}[f,g] = \sum_{i=1}^n \mathbb{E}[c_i d_i]$. Let

$$c_{i,0} = \mathbb{E}[f_0|(x_1,\dots,x_i)] - \mathbb{E}[f_0|(x_1,\dots,x_{i-1})]\,,$$

for $1 \le i \le n-1$, and define $c_{i,1}$, $d_{i,0}$, and $d_{i,1}$ similarly. Then we have $c_i = (c_{i,0}+c_{i,1})/2$, $d_i = (d_{i,0}+d_{i,1})/2$, and we can write the covariance as:

$$\mathrm{Cov}[f,g] = \sum_{i=1}^{n} \mathbb{E}[c_i d_i] = \frac{1}{4}\sum_{i=1}^{n-1}\sum_{a,b\in\{0,1\}} \mathbb{E}[c_{i,a} d_{i,b}] + \mathbb{E}[c_n d_n] = \frac{1}{4}\sum_{a,b\in\{0,1\}} \mathrm{Cov}[f_a,g_b] + \mathbb{E}[c_n d_n]\,.$$

By the triangle inequality, $|\mathrm{Cov}[f,g]| \le (1/4)\sum_{a,b\in\{0,1\}}|\mathrm{Cov}[f_a,g_b]| + |\mathbb{E}[c_n d_n]|\,.$

Since f_a and g_b are functions on $n-1$ variables we can use the induction hypothesis, and we have:

$$|\mathrm{Cov}[f,g]| \leq \frac{1}{4} \sum_{a,b\in\{0,1\}} \sum_{i=1}^{n-1} \delta_i(T_a)\,\mathrm{Inf}_i(g_b) + |\mathbb{E}[c_n d_n]| = \sum_{i=1}^{n-1} \delta_i(T)\,\mathrm{Inf}_i(g) + |\mathbb{E}[c_n d_n]|\,.$$

As $\mathrm{Inf}_n(g) = \mathrm{Inf}_n(-g)$, we have that $|\mathbb{E}[c_n d_n]| \leq \mathrm{Inf}_n(g)$ by Lemma 2.2, and $\delta_n(T) = 1$ because x_n is the root of the tree. Thus, the inductive step holds. □

References

[1] MICHAEL BEN-OR AND NATHAN LINIAL: Collective coin flipping. In S. MICALI, editor, *Randomness and Computation*, pp. 91–115. Acad. Press, 1989. Preliminary version in *Proc. FOCS'85*. 82

[2] BRADLEY EFRON AND CHARLES STEIN: The jackknife estimate of variance. *Ann. Statist.*, 9:586–596, 1981. [doi:10.1214/aos/1176345462]. 82

[3] RYAN O'DONNELL, MICHAEL SAKS, ODED SCHRAMM, AND ROCCO A. SERVEDIO: Every decision tree has an influential variable. In *Proc. 46th FOCS*, pp. 31–39. IEEE Computer Soc., 2005. [doi:10.1109/SFCS.2005.34]. 82

[4] RYAN O'DONNELL AND ROCCO A. SERVEDIO: Learning monotone decision trees in polynomial time. *SIAM J. Comput.*, 37(3):827–844, 2007. Preliminary version in *Proc. of CCC'06*. [doi:10.1137/060669309]. 81

[5] J. MICHAEL STEELE: An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.*, 14:753–758, 1986. [doi:10.1214/aos/1176349952]. 82

AUTHOR

Homin K. Lee
Department of Computer Science
University of Texas at Austin
homin@cs.utexas.edu
http://www.cs.utexas.edu/~homin

ABOUT THE AUTHOR

HOMIN K. LEE received his Ph. D. from the Department of Computer Science at Columbia University in 2009 under the supervision of Tal Malkin and Rocco A. Servedio. Homin is currently a postdoctoral researcher at the University of Texas at Austin. Homin's research interests are in computational learning theory, combinatorics, computational complexity, and cryptography. Homin is also currently hungry.