
Accelerating Decision Trees with SIMD

Tara Mirmira and Ramesh Subramonian

September 21, 2021

Abstract

In this paper we show the benefits and limitations of using SIMD techniques to accelerate decision trees, in both the building phase and the inference phase.

1 Introduction

TO BE COMPLETED

1.1 ISPC

We chose to use ISPC for “SIMD-fication”. The Intel® Implicit SPMD Program Compiler (Intel® ISPC) is a compiler for writing SPMD (single program multiple data) programs to run on the CPU and GPU

Our reasons for doing so are best stated by the documentation itself

1. *Build a variant of the C programming language that delivers good performance to performance-oriented programmers who want to run SPMD programs on CPUs and GPUs.*
2. *Provide a thin abstraction layer between the programmer and the hardware—in particular, to follow the lesson from C for serial programs of having an execution and data model where the programmer can cleanly reason about the mapping of their source program to compiled assembly language and the underlying hardware.*
3. *Harness the computational power of the Single Program, Multiple Data (SIMD) vector units without the extremely low-programmer-productivity activity of directly writing intrinsics.*
4. *Explore opportunities from close-coupling between C/C++ application code and SPMD ispc code running on the same processor—lightweight function calls between the two languages, sharing data directly via pointers without copying or reformatting, etc.*

1.2 Inferencing

Given a decision tree, T , and a set of data points $X = \{x_1, x_2, \dots\}$, the goal is to create $Y = \{y_1, y_2, \dots\}$ where y_i is the leaf to which x_i is assigned.

1.2.1 Random Forests