Parallelization of RSVP Dataset Generator

Nick Carey

May 9, 2014

Abstract

Brain Computer Interfaces have the potential to help interactively explore and classify large datasets. Rapid Serial Visual Presentations, or RSVP, is a BCI method that must be further tested on visual datasets in order to find novel applications. We have developed a program that generates abstract visual datasets on the fly in order to test the capabilities of RSVP. In order to speed up dataset generation to match the pace demanded by RSVP, we analyze and implement parallelization of the data generation program. Using Amdahl's Law, we find the maximum speedup possible. We compare actual results with the results predicted by Amdahl's Law and reason about causes for any non-optimal speedup.

1 Introduction

Brain Computer Interfaces (BCI) have the potential to become a new tool for the data scientist. By reading brain signals rather than mouse and keyboard signals, the loop between computer and human can be considerably shortened. Rapid Serial Visual Presentations (RSVP) is one such BCI system that reads human subject brain signals while rapidly presenting the subject with approximately ten images per second. The particular RSVP system in question will look for the P300 brain wave. The P300 brain wave is emitted whenever the subject sees something interesting. Therefore, by associating specific images with the P300 wave, images are classified as 'interesting' or 'non-interesting'.

The meaning of the P300 wave can be quite vague. 'Interesting' is a very situational adjective. Therefore it is important to investigate which sort of datasets and data visualizations could be classified using a P300 wave. In order to better understand the capabilities of RSVP, we have iteratively developed and tested an abstract dataset. Our experimental goal with RSVP is to see if we can tease out a rare, low-dimensional signal hidden in a high-dimensional noisy dataset; finding hidden low-dimensional signal is common to some astronomy and cyber forensics problems.

We have had success using RSVP to classify views of the abstract dataset that show some signal versus images showing only noise. However, we currently prepare decks of images before running RSVP; in the future we would like to give RSVP online views of the abstract dataset based on previous RSVP results. For example, if RSVP triggers on an image showing some of the hidden structure, then in the next iteration of RSVP we would like to show similar views of the dataset in order to drill down completely on the hidden signal. However, RSVP is very fast and our abstract dataset query program is very slow. In order to match the pace demanded by RSVP, our query program must be able to execute and visualize at least ten queries per second.

In section 2, we outline the abstract dataset generator and query program and analyze which program sections are parallelizable. Using Amdahl's Law[1] we calculate optimal speedup. In section 3, we describe an experiment to test our parallel implementation, and in section 4 we show the results of the experiment. In section 5 we analyze the results and discuss any causes for non-optimal speedup.

2 Dataset Generator Program

The abstract dataset generator program we have developed to test the capabilities of RSVP consists of several steps. In step one, we create a set of N 3-dimensional coordinates that form a wire-frame cube. We then add on a few more dimensions of uniform random noise. After step one, we have a starting dataset of six dimensions, three of which form coordinates for a wire-frame cube and three of which are random noise.

In step two, we apply a random six-dimensional rotation to our starting dataset. After the random rotation, we have our simulated dataset. Visualizations of the simulated dataset should only show noise; we have hidden our low-dimensional cube signal inside a high dimensional noisy dataset. The simulated dataset is what models some astronomy and cyber datasets.

In step three, we query the simulated dataset approximately 1000 times. Each query is another random six-dimensional rotation applied to the simulated dataset. Some rare random rotations will be 'lucky' and close to the opposite of the original six-dimensional rotation used to create the simulated dataset; visualizations of these 'lucky' rotations will show some of the hidden cubic signal structure while most random rotations will only show noise.

In step four, we visualize the random rotations applied to the simulated dataset to create images appropriate for use with RSVP.

2.1 Amdahl's Law Analysis

lolol[1]

3 Experimental Method and Setup

4 Results

5 Discussion

References

[1] Gene M. Amdahl. Validity of the single processor approach to achieving large scale computing capabilities. In *Proceedings of the April 18-20, 1967, Spring Joint Computer Conference*, AFIPS '67 (Spring), pages 483–485, New York, NY, USA, 1967. ACM.