GPU Data Mining in Neuroimaging Genomics

Bob Zigon

Beckman Coulter
Indianapolis, Indiana

May 10, 2017
Outline

- Background
- ANOVA for Voxels and SNPs
- VEGAS for Voxels and Genes
- High Speed GPU Monte-Carlo Simulator
What is Neuroimaging Genomics?

1. Neuroimaging Genomics is the fusion of brain imaging and genotyping data.

2. Study the influence of genetic variation on brain structure and function.
MRI and Sequencing data

MRI instrument

MRI data

Genotyping instrument

Genotyping data
Develop an interactive tool for studying Alzheimer’s Disease by coupling a 3D brain explorer with a genome explorer.

Prior Art
120 ROI’s ⇒ 1,000,000 voxels
20,000 SNP’s

Our Goal
1,000,000 voxels
1,000,000 SNP’s
Problem Definition

Brain with 120 Regions of Interest

Brain with 1,000,000 voxels
Brain Explorer
How We Do It – The UI

SNP Explorer

Brain Explorer
How We Do It – The UI

SNP Explorer

Brain Explorer

Heat Map
ANOVA - Analysis of Variance

Understand the relationship between the gray matter density from the MRI and the SNP genotype.

"Did the combination happen by chance or not?"
ANOVA - Analysis of Variance

Understand the relationship between the gray matter density from the MRI and the SNP genotype.

"Did the combination happen by chance or not?"

Computational complexity $O(N_v \times N_j \times N_s)$

$N_v$ - number of voxels
$N_j$ - number of subjects
$N_s$ - number of SNPS
ANOVA

\[
\begin{bmatrix}
v_{11} & \cdots & v_{1M} \\
v_{21} & \cdots & v_{2M} \\
\vdots & \ddots & \vdots \\
v_{N1} & \cdots & v_{NM}
\end{bmatrix}
\]

ANOVA

\[
\begin{bmatrix}
s_{11} & \cdots & s_{1M} \\
s_{21} & \cdots & s_{2M} \\
\vdots & \ddots & \vdots \\
s_{K1} & \cdots & s_{KM}
\end{bmatrix}
\]

ANOVA

\[
\begin{bmatrix}
vs_{11} & \cdots & vs_{1M} \\
vs_{21} & \cdots & vs_{2M} \\
\vdots & \ddots & \vdots \\
vs_{N1} & \cdots & vs_{NM}
\end{bmatrix}
\]
VEGAS - VErsatile Gene based Association Study

Understand the relationship between the gray matter density from the MRI and the collective effect of multiple SNPs within a gene.

"Did the combination happen by chance or not?"
VEGAS - VErsatile Gene based Association Study

Understand the relationship between the gray matter density from the MRI and the collective effect of multiple SNPs within a gene.

"Did the combination happen by chance or not?"

Computational complexity $\mathcal{O}(N_v \times N_j \times N_s + N_v \times N_g \times N_i)$

- $N_v$ - number of voxels
- $N_g$ - number of genes
- $N_i$ - number of Monte-Carlo iterations ($10^2, 10^3, 10^4, 10^5$, or $10^6$)
VEGAS

\[
\begin{bmatrix}
    v_{11} & \cdots & v_{1M} \\
    v_{21} & \cdots & v_{2M} \\
    \vdots & \ddots & \vdots \\
    v_{N1} & \cdots & v_{NM}
\end{bmatrix}
\]

\[
\begin{bmatrix}
    s_{11} & \cdots & s_{1M} \\
    s_{21} & \cdots & s_{2M} \\
    \vdots & \ddots & \vdots \\
    s_{K1} & \cdots & s_{KM}
\end{bmatrix}
\]

\[
\begin{bmatrix}
    v_{s11} & \cdots & v_{s1M} \\
    v_{s21} & \cdots & v_{s2M} \\
    \vdots & \ddots & \vdots \\
    v_{sN1} & \cdots & v_{sNM}
\end{bmatrix}
\]

\[
\begin{bmatrix}
    v_{g11} & \cdots & v_{g1G} \\
    v_{g21} & \cdots & v_{g2G} \\
    \vdots & \ddots & \vdots \\
    v_{gN1} & \cdots & v_{gNG}
\end{bmatrix}
\]

ANOVA

Monte Carlo Simulation
Video of the Interactive Neuroimaging Genomic Browser
Lessons Learned

How do you build a high speed Monte-Carlo Simulator for an N dimensional problem on a GPU?
One Dimensional

for $i = 1$ to $K$ do
    Choose $X$ from $N(0,1)$
    $Y = F(X)$
    Make decision about $Y$
end
One Dimensional

\[
\text{for } i = 1 \text{ to } K \text{ do }
\]
\[
\begin{align*}
\text{Choose } X & \text{ from } \mathcal{N}(0,1) \\
Y &= F(X) \\
\text{Make decision about } Y
\end{align*}
\]
\text{end}

N-Dimensional

\[
\text{for } i = 1 \text{ to } K \text{ do }
\]
\[
\begin{align*}
\text{Choose } n \text{ values from } \mathcal{N}(0,1) \text{ giving } X^n \\
Y^n &= F(X^n) \\
\text{Make decision about } Y^n
\end{align*}
\]
\text{end}
Lessons Learned

First Attempt at N-Dimensional

```
foreach voxel V in parallel do
    foreach gene G in parallel do
        for i = 1 to K do
            Choose n values from N(0,1) giving $X^n$
            $Y^n = F(X^n)$
            Make decision about $Y^n$
        end
    end
end
```
Lessons Learned

First Attempt at N-Dimensional

```plaintext
foreach voxel V in parallel do
  foreach gene G in parallel do
    for i = 1 to K do
      Choose n values from N(0,1) giving $X^n$
      $Y^n = F(X^n)$
      Make decision about $Y^n$
    end
  end
end
```

<table>
<thead>
<tr>
<th></th>
<th>Slow</th>
<th>Fast</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Bandwidth (gb/sec)</td>
<td>20</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>GFLOPS</td>
<td>20</td>
<td>10,000</td>
<td></td>
</tr>
</tbody>
</table>
N-Dimensional Ah-Ha

In parallel, generate \( n \times K \) values from \( \mathcal{N}(0,1) \) giving \( X^{n \times K} \)

\[
Y^{n \times K} = F^{n \times n} \times X^{n \times K}
\]

In parallel, decide about \( Y^{n \times K} \)
Lessons Learned

Slow N-Dimensional

```
foreach voxel V in parallel do
  foreach gene G in parallel do
    for i = 1 to K do
      n values from N(0,1)
      \(Y^n = F(X^n)\)
      Make decision about \(Y^n\)
    end
  end
end
```

Fast N-Dimensional

```
foreach voxel V sequentially do
  foreach gene G sequentially do
    In parallel, generate nK values
    \(Y^{n\times K} = F^{n\times n} \times X^{n\times K}\)
    In parallel, decide about \(Y^{n\times K}\)
  end
end
```
Lessons Learned

Slow N-Dimensional

\[
\text{foreach voxel } V \text{ in parallel do}
\]
\[
\quad \text{foreach gene } G \text{ in parallel do}
\]
\[
\quad\quad \text{for } i = 1 \text{ to } K \text{ do}
\]
\[
\quad\quad\quad \text{n values from } \mathcal{N}(0,1)
\]
\[
\quad\quad\quad Y^n = F(X^n)
\]
\[
\quad\quad \text{Make decision about } Y^n
\]
\[
\quad \text{end}
\]
\[
\text{end}
\]
\[
\text{end}
\]

Fast N-Dimensional

\[
\text{foreach voxel } V \text{ sequentially do}
\]
\[
\quad \text{foreach gene } G \text{ sequentially do}
\]
\[
\quad\quad \text{In parallel, generate } nK \text{ values}
\]
\[
\quad\quad Y^{n \times K} = F^{n \times n} \times X^{n \times K}
\]
\[
\quad \text{In parallel, decide about } Y^{n \times K}
\]
\[
\quad \text{end}
\]
\[
\text{end}
\]
Lessons Learned

Slow N-Dimensional

```plaintext
foreach voxel V in parallel do
    foreach gene G in parallel do
        for $i = 1$ to $K$ do
            n values from $N(0,1)$
            $Y^n = F(X^n)$
            Make decision about $Y^n$
        end
    end
end
```

Fast N-Dimensional

```plaintext
foreach voxel V sequentially do
    foreach gene G sequentially do
        In parallel, generate $nK$ values
        $Y^{n \times K} = F^{n \times n} \times X^{n \times K}$
        In parallel, decide about $Y^{n \times K}$
    end
end
```

<table>
<thead>
<tr>
<th></th>
<th>Slow</th>
<th>Fast</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Bandwidth (gb/sec)</td>
<td>20</td>
<td>155</td>
<td>320</td>
</tr>
<tr>
<td>GFLOPS</td>
<td>20</td>
<td>2,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

800X improvement!
Figure 1: Execution times for 1 VEGAS run with K=10,000 Monte-Carlo iterations
Acknowledgements

1. Professor Li Shen, Dept. of Radiology and Imaging Sciences, IU School of Medicine, NIH R01 EB022574, R01 LM011360, U01 AG024904, and IUPUI ITDP Program.

2. Professor Shiaofen Fang, Computer Science Dept. Chair, Indiana University-Purdue University of Indianapolis

3. Professor Mohammad Al Hasan, Associate Professor Computer Science, Indiana University-Purdue University of Indianapolis

4. Huang Li, PhD Candidate, Computer Science, Indiana University-Purdue University of Indianapolis
Thank you

robert.zigon@beckman.com