Powerset Viewer: A Datamining Application

Update

Completed Tools and Features
- And relevant GUI widgets
- Implemented animation between zoom states and automatic zooming
- Increased alphabet size from 14 to 30
  - Optimized calculations

Completed Tools and Features
- And relevant GUI widgets
- Implemented animation between zoom states and automatic zooming
- Increased alphabet size from 14 to 30
- Optimized calculations
- Realized set cardinality is, in practice, low
- Using max set size of 10
Milestones Status Update

- #1 Completion of the basic visualization of a randomized database of small set size (~10)
- #2 Addition of a single level of “marking”.
- #3 Addition of multiple levels of “marking” (6)
- #4 Addition of background marking to demarcate areas of sets containing different amounts of items.
- #5 Implement multiple constraints
- #6 Increase maximum possible dataset size to at least 100.

Difficulties

- BigInteger solution to increase maximum alphabet caused massive slow-down
  - Recall: required BigIntegers to support > 30 alphabet size
  - Solution: redesign keys to use integers and create a bridge to map integers to BigInteger positions

BEFORE BRIDGE

- Incoming Set (Position = 982) Success!
- Incoming Set (Position = 2^32 + 1) CRASH!
  - Integer too large
AFTER BRIDGE

- Incoming Set (Position = 982)
  - Encode to Key #1 Success!
- Incoming Set (Position = 2^32 + 1)
  - Encode to Key #2 Success!
- Incoming Set (Position = arbitrarily large)
  - Encode to Key #3 Success!

Difficulties

- BigInteger solution to increase maximum alphabet caused massive slow-down
  - Recall: required BigIntegers to support > 30 alphabet size
  - Solution: redesign keys to use integers and create a bridge to map integers to BigInteger positions
- Expensive initial costs
- Grid size limited by integer restrictions
  - Solution: create grid on the fly

Benchmarks

- Low Cardinality First

<table>
<thead>
<tr>
<th>MEMORY (MB)</th>
<th>SET COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>10M</td>
</tr>
<tr>
<td>75</td>
<td>1M</td>
</tr>
<tr>
<td>74</td>
<td>100,000</td>
</tr>
<tr>
<td>73</td>
<td>10,000</td>
</tr>
<tr>
<td>58</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Figure: Low Cardinality (10000 sets) 73 MB

Benchmarks (cont’d)

- Random Generated

<table>
<thead>
<tr>
<th>MEMORY (MB)</th>
<th>SET COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>263</td>
</tr>
<tr>
<td>71</td>
<td>168</td>
</tr>
<tr>
<td>70</td>
<td>127</td>
</tr>
<tr>
<td>72</td>
<td>30</td>
</tr>
<tr>
<td>71</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure: Random (176 sets) 71 MB
Questions and Comments