Progressive Visual Analytics
User-Driven Visual Exploration of In-Progress Analytics

Chad Stolper
Georgia Tech

Adam Perer
IBM T.J. Watson Research Center

David Gotz
UNC Chapel Hill
Visual Analytics Workflow
Visual Analytics Workflow

Select Dataset
Visual Analytics Workflow
Visual Analytics Workflow

1. Select Dataset
2. Select Analytic Parameters
3. Run Analytic
Visual Analytics Workflow

1. Select Dataset
2. Select Analytic Parameters
3. Run Analytic
4. Visualize Results
Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Visualize Results → Interpret Results
Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Visualize Results → Interpret Results

Select Analytic Parameters"
Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Wait for Analytic to Complete → Visualize Results → Interpret Results
Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Wait for Analytic to Complete → Visualize Results → Interpret Results

Bad!
Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Wait for Analytic to Complete → Visualize Results → Interpret Results

Bad!
Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Wait for Analytic to Complete → Visualize Results → Interpret Results

Very Bad!
1. Increasingly large quantities of data

https://www.flickr.com/photos/veggiefrog/3435380297/
1. Increasingly large quantities of data
2. Increasingly complex analytics

https://www.flickr.com/photos/veggiefrog/3435380297/
1. Increasingly large quantities of data

2. Increasingly complex analytics

https://www.flickr.com/photos/veggiefrog/3435380297/
Batch Visual Analytics Workflow

Select Dataset → Select Analytic Parameters → Run Analytic → Wait for Analytic to Complete → Visualize Results → Interpret Results
Batch Visual Analytics Workflow
Batch Visual Analytics Workflow

1. Select Dataset
2. Select Analytic Parameters
3. Run Analytic
4. Visualize Partial Results
5. Interpret Partial Results
6. Visualize Complete Results
7. Interpret Complete Results
Batch Visual Analytics Workflow

1. Select Dataset
2. Select Analytic Parameters
3. Run Analytic
4. Visualize Partial Results
5. Interpret Partial Results
6. Visualize Complete Results
7. Interpret Complete Results
Batch Visual Analytics Workflow
Batch Visual Analytics Workflow

Progressive
The "filter" box allows an analyst to create a filter on either a dimension or measures. In Figure 1, for example, the analyst performs a query over an FAA (US Federal Aviation Administration) database of flight delays, showing the average arrival delay by day of the week.

Visualization of Queries

When a query is issued, the system sends it to the back-end, which computes and returns an estimate and confidence bounds; the front-end displays a chart of the results. The estimate and bounds are updated every second with more rows of data.

When is this strategy viable?

How do we design systems to take advantage of it?
Progressive Visual Analytics

Progressive, User-Driven Analytics
Progressive, Interactive Visualization
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Progressive Analytics

Semantically-Meaningful Partial Results
Progressive Analytics

Semantically-Meaningful Partial Results
(Thomas-- Results Feedback)
Progressive Analytics

Semantically-Meaningful Partial Results

The partial results are of the same form as the final results
Progressive Analytics

Semantically-Meaningful Partial Results

Progressive Analytics

Semantically-Meaningful Partial Results

Progressive Analytics

Semantically-Meaningful Partial Results

Progressive Analytics

Semantically-Meaningful Partial Results

Progressive Analytics

Semantically-Meaningful Partial Results

K-Means
Progressive Analytics

Semantically-Meaningful Partial Results

K-Means Logistic Regression
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
User-Driven Analytics

Incorporate Analyst Knowledge
User-Driven Analytics

Incorporate Analyst Knowledge
User-Driven Analytics

Incorporate Analyst Knowledge
User-Driven Analytics

Incorporate Analyst Knowledge

(Thomas– Result Control)
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Progressive Visualization

Visualize the Partial Results
Progressive Visualization

Visualize the Partial Results
(Thomas– Results Feedback)
Progressive Visualization

Analytic Visualization Analyst
Progressive Visualization

Up-to-Date Information

Analytic Visualization Analyst
Progressive Visualization

Up-to-Date Information

Analytic Visualization Analyst
Progressive Visualization

Up-to-Date Information

Analytic Visualization Analyst

http://info.jmu.edu/dux/files/2013/01/chaos.jpg
Progressive Visualization

Up-to-Date Information

Analytic Visualization Analyst

http://info.jmu.edu/dux/files/2013/01/chaos.jpg
http://www.historyforkids.org/learn/medieval/history/byzantine/justinian.jpg
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Progressive Visualization

Analytic  Visualization  Analyst

Up-to-Date Information
Interactive Visualization

Analytic -> Visualization -> Analyst

Up-to-Date Information

Analyst’s Domain Knowledge
Interactive Visualization

Analytic \quad Visualization \quad Analyst

Up-to-Date Information

Analyst’s Domain Knowledge

(Thomas–Execution Control)
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
So what might such a system look like?
Electronic Medical Records
Electronic Medical Records

Procedures
Lab Tests
Diagnoses
Medications
Electronic Medical Records

Event Series
Electronic Medical Records

Event Series
Electronic Medical Records

Frequent Patterns

Event Series
Electronic Medical Records

Frequent Patterns

Event Series
Electronic Medical Records

Frequent Patterns
Event Series
Correlate with Outcomes
Electronic Medical Records

What Works?

Frequent Patterns
Event Series
Correlate with Outcomes
Electronic Medical Records

What Works?
(What Doesn’t Work?)

Frequent Patterns
Event Series
Correlate with Outcomes
Sequential Pattern Mining Algorithm

Sequential Pattern Mining Algorithm
(Thomas– Dependent Subdivision)

Sequential **Pattern** Mining Algorithm

- Build a tree of every possible pattern
  - (up to a maximum length)

- Prune the tree
  - (using a support threshold)
Sequential Pattern Mining Algorithm

- Build a tree of every possible pattern
  - (up to a maximum length)
- Prune the tree
  - (using a support threshold)
Sequential Pattern Mining Algorithm

- Build a tree of every possible pattern
  - (up to a maximum length)

- Prune the tree
  - (using a support threshold)
Sequential Pattern Mining Algorithm

- Build a tree of every possible pattern
  - (up to a maximum length)
- Prune the tree
  - (using a support threshold)
Sequential Pattern Mining Algorithm

- Frequent Patterns
- Event Series
- Correlate with Outcomes

- Depth-First Search
- Outputs frequent patterns as each is found
Sequential Pattern Mining Algorithm

Frequent Patterns
Event Series
Correlate with Outcomes

- Depth-First Search
- Outputs frequent patterns as each is found
Sequential Pattern Mining Algorithm

- Frequent Patterns
- Event Series
- Correlate with Outcomes

- Depth-First Search
- Outputs frequent patterns as each is found
Interactive Sequential Pattern Mining

- Breadth-First Search
  - Outputs frequent patterns as each is found

Frequent Patterns
Event Series
Correlate with Outcomes
Interactive Sequential Pattern Mining

- Breadth-First Search
  - Outputs frequent patterns as each is found

Frequent Patterns
Event Series
Correlate with Outcomes

HF  BB  D  COPD  BB  D
Interactive Sequential Pattern Mining

- Frequent Patterns
- Event Series
- Correlate with Outcomes
- Breadth-First Search
  - Shorter patterns first
- Outputs frequent patterns as each is found
Interactive Sequential Pattern Mining

- Breadth-First Search
  - Shorter patterns first

- Outputs frequent patterns as each is found

http://thenextweb.com/wp-content/blogs.dir/1/files/2013/02/queue.jpg
Interactive Sequential Pattern Mining

- Breadth-First Search
  - Shorter patterns first
  - Analyst prioritization
- Outputs frequent patterns as each is found

http://thenextweb.com/wp-content/blogs.dir/1/files/2013/02/queue.jpg
Interactive Sequential Pattern Mining

- Breadth-First Search
  - Shorter patterns first
  - Analyst prioritization
- Outputs frequent patterns as each is found
Interactive **Sequential Pattern Mining**

- **Breadth-First Search**
  - Shorter patterns first
  - Analyst prioritization

- Outputs frequent patterns as each is found

- **Manual Pruning**

---

[Image: http://www.bluestonegarden.com/media/catalog/product/cache/1/image/9df78eab33525d08d6e5fb8d27136e95/WOLF-Garten-Telescoping-Hedge-Shears-1.png]
Interactive Sequential Pattern Mining

- Breadth-First Search
  - Shorter patterns first
  - Analyst prioritization

- Outputs frequent patterns as each is found

- Manual Pruning
Interactive **Sequential Pattern Mining**

- Breadth-First Search
  - Shorter patterns first
  - Analyst prioritization
- Outputs frequent patterns as each is found
- Manual Pruning
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Progressive Visual Analytics

Progressive, User-Driven Analytics

Progressive, Interactive Visualization
Selected Pattern: heart failure -> chronic airways obstruction, not elsewhere classified -> diuretics3
Support: 140 (32.56%)
Correlation: 0.1022
AdjCorrelation: 0.0859

Pause and restart using only patients with this pattern
<table>
<thead>
<tr>
<th>Support</th>
<th>[Correlation]</th>
</tr>
</thead>
<tbody>
<tr>
<td>heart failure</td>
<td>angiotensin-converting enzyme (ace) inhibitors</td>
</tr>
<tr>
<td>diuretics3</td>
<td>diuretics3---&gt;heart failure---&gt;diuretics3</td>
</tr>
<tr>
<td>chronic airways obstruction, not elsewhere classified</td>
<td>heart failure---&gt;angiotensin-converting enzyme (ace) inhibitors</td>
</tr>
<tr>
<td>heart failure---&gt;diuretics3</td>
<td>diuretics3---&gt;heart failure---&gt;glucocorticoids---&gt;diuretics3</td>
</tr>
<tr>
<td>heart failure---&gt;chronic airways obstruction, not elsewhere classified</td>
<td>dyslipidemics, hmg coa reductase inhibitors</td>
</tr>
<tr>
<td>diuretics3---&gt;heart failure</td>
<td>beta blockers2</td>
</tr>
<tr>
<td>beta blockers2</td>
<td>diuretics3---&gt;heart failure---&gt;diuretics3---&gt;chronic airways obstruction</td>
</tr>
<tr>
<td>glucocorticoids</td>
<td>diuretics3</td>
</tr>
<tr>
<td>chronic airways obstruction, not elsewhere classified---&gt;heart failure</td>
<td>dyslipidemics, hmg coa reductase inhibitors---&gt;angiotensin-converting enzyme</td>
</tr>
<tr>
<td>heart failure---&gt;beta blockers2</td>
<td>diuretics3---&gt;heart failure---&gt;chronic airways obstruction, not elsewhere classified</td>
</tr>
</tbody>
</table>
heart failure
- diuretics3
- chronic airways obstruction, not elsewhere classified
- heart failure--diuretics3
- heart failure--chronic airways obstruction, not elsewhere classified
- diuretics3--heart failure
- beta blockers2
- glucocorticoids
- chronic airways obstruction, not elsewhere classified--heart failure1
- heart failure--beta blockers2

angiotensin-converting enzyme (ace) inhibitors
- diuretics3-->heart failure-->diuretics3
- heart failure-->angiotensin-converting enzyme (ace) inhibitors
- diuretics3-->heart failure-->glucocorticoids-->diuretics3
- dyslipidemics, hmg coa reductase inhibitors
- beta blockers2
- diuretics3-->heart failure-->diuretics3-->chronic airways obstruction
- diuretics3
- dyslipidemics, hmg coa reductase inhibitors--angiotensin-converting enzyme (ace) inhibitors
- diuretics3-->heart failure-->chronic airways obstruction, not elsewhere classified
<table>
<thead>
<tr>
<th>Support</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>heart failure</td>
<td>angiotensin-converting enzyme (ace) inhibitors</td>
</tr>
<tr>
<td>diuretics3</td>
<td>diuretics3---&gt;heart failure---&gt;diuretics3</td>
</tr>
<tr>
<td>chronic airways obstruction, not elsewhere classified</td>
<td>heart failure---&gt;angiotensin-converting enzyme (ace) inhibitors</td>
</tr>
<tr>
<td>heart failure---&gt;diuretics3</td>
<td>diuretics3---&gt;heart failure---&gt;glucocorticoids---&gt;diuretics3</td>
</tr>
<tr>
<td>heart failure---&gt;chronic airways obstruction, not elsewhere classified</td>
<td>dyslipidemics, hmg coa reductase inhibitors</td>
</tr>
<tr>
<td>diuretics3---&gt;heart failure</td>
<td>beta blockers2</td>
</tr>
<tr>
<td>beta blockers2</td>
<td>diuretics3---&gt;heart failure---&gt;diuretics3---&gt;chronic airways obstruction, not elsewhere classified</td>
</tr>
<tr>
<td>glucocorticoids</td>
<td>diuretics3---&gt;heart failure---&gt;glucocorticoids</td>
</tr>
<tr>
<td>chronic airways obstruction, not elsewhere classified---&gt;heart failure</td>
<td>diuretics3</td>
</tr>
<tr>
<td>heart failure---&gt;beta blockers2</td>
<td>dyslipidemics, hmg coa reductase inhibitors---&gt;angiotensin-converting enzyme (ace) inhibitors</td>
</tr>
<tr>
<td>diuretics3</td>
<td>diuretics3---&gt;heart failure---&gt;chronic airways obstruction, not elsewhere classified</td>
</tr>
</tbody>
</table>
Selected Pattern: heart failure -> chronic airways obstruction, not elsewhere classified -> diuretics3
Support: 140 (32.56%)
Correlation: 0.1022
ΔCorrelation: 0.059

Minimum Support Required: 64/430 (15%)
Level of Detail: 0
Maximum Length: 6
Positive Outcome Indicator: 1

Pause and restart using only patients with this pattern.
So Does it Work?
So Does it Work?

Medical Researchers,
University Hospital of North Norway
So Does it Work?

AMIA 2014
Skrovseth, Perer, Delaney, Revaug, Lindsetmo, Augestad

"Detecting Novel Associations for Surgical Hospital Readmissions in Large Datasets by Interactive Visual Analytics"
Contributions

- A precise definition and design guidelines for Progressive Visual Analytics
- An example Progressive Visual Analytics system, Progressive Insights
- (In the paper) A case study of using Progressive Insights for exploring frequent patterns in EHR
Contributions

- A precise definition and design guidelines for Progressive Visual Analytics
- An example Progressive Visual Analytics system, Progressive Insights
- (In the paper) A case study of using Progressive Insights for exploring frequent patterns in EHR
Contributions

- A precise definition and design guidelines for Progressive Visual Analytics
- An example Progressive Visual Analytics system, Progressive Insights
- (In the paper) A case study of using Progressive Insights for exploring frequent patterns in EHR
Contributions

- A precise definition and design guidelines for Progressive Visual Analytics
- An example Progressive Visual Analytics system, Progressive Insights
- (In the paper) A case study of using Progressive Insights for exploring frequent patterns in EHR
Chad Stolper
Georgia Tech
chadstolper@gatech.edu

Adam Perer
IBM T.J. Watson Research Center

David Gotz
UNC Chapel Hill
Thank You!

Chad Stolper
Georgia Tech
chadstolper@gatech.edu

Adam Perer
IBM T.J. Watson Research Center

David Gotz
UNC Chapel Hill
Questions?

Chad Stolper
Georgia Tech
chadstolper@gatech.edu

Adam Perer
IBM T.J. Watson Research Center

David Gotz
UNC Chapel Hill
Questions?

Chad Stolper
Georgia Tech
chadstolper@gatech.edu

Adam Perer
IBM T.J. Watson Research Center

David Gotz
UNC Chapel Hill
Progressive Visual Analytics

Seven Design Guidelines
Progressive Visual Analytics Systems

Analytics should…

1. Provide increasingly meaningful partial results as the algorithm executes

2. Allow users to focus the algorithm to subspaces of interest

3. Allow users to ignore irrelevant subspaces
Progressive Visual Analytics Systems

Visualizations should…

4. Minimize distractions by not changing views excessively

5. Provide cues to indicate where new results have been found

6. Support on-demand refresh

7. Provide an interface to specify where the analytics should focus and ignore