An Introduction to Information Visualization Techniques for Exploring Large Database

Jing Yang
Fall 2005

Evaluation in Information Visualization

Class 3
Motivation

- What are the advantages and limitations of this technique (system)?
- Are there any improvement of this new technique over the existing ones?
- Is this technique (system) just “cool” or actually helps people?

Measures

- Different possible ways to measure:
  - Impact on community as a whole, influential ideas
  - Assistance to people in the tasks they care about

- Dr. John Stasko, Slides of CS7500 at Gatech
Strong View

- Unless a new technique or tool helps people in some kind of problem or task, it doesn’t have any value

- Dr. John Stasko, Slides of CS7500 at Gatech

Broaden Thinking

- Sometimes the chain of influence can be long and drawn out
  - System X influences System Y influences System Z which is incorporated into a practical tool that is of true value to people
  - This is what research is all about (typically)

- Dr. John Stasko, Slides of CS7500 at Gatech
Survey of Evaluation

- I read through Infovis 95- Infovis 2001 for papers that included some form of evaluation.
Evaluation In InfoVis

- Evaluation has recently been regarded as important.
- Many evaluations are ad hoc.
  - Ad hoc questionnaires.
  - Ad hoc processes.
  - Insufficient subjects.
- In HCI field, there is a long history of evaluation and systematic theory and approaches

Evaluating UI vs. InfoVis

- Difference
  - Usability vs. Utility
    - Imagine a visualization that are very usable but not useful
    - More domain knowledge and situated use are required
Evaluating InfoVis in General

- Very difficult in InfoVis to compare “apples to apples”
  - Hard to compare System A to System B
  - Different tools were built to address different user tasks
- UI can heavily influence utility and value of visualization technique
- Borrow UI evaluation methods, adapt it to InfoVis

Usability Inspection in UI

- Usability inspection - generic name for methods based on having evaluators inspect or examine usability-related aspects of a user interface [NL94]
- General approaches
  - user testing, heuristic evaluation, feature inspection, heuristic estimation, consistency inspection, standards inspection, pluralistic walkthrough, and cognitive walkthrough [Nie95]
What to Measure: Usability Metrics

Three categories:
- Preference metrics
- Performance metrics
- Predictive metrics (design metrics)

Preference metrics

- Subjective evaluations and preferences of users
- Questionnaires and rating scales are used.
- Standard forms or questionnaires are more reliable than ad hoc surveys
- Approaches:
  - User testing
  - Heuristic evaluation
Performance Metrics

- Actual use of working software
  - Example: How fast...? How many...?
- Approach
  - User testing

Design Metrics

- Quality of designs or prototypes.
- They can be measured without working systems!
- Examples:
  - Essential efficiency (brief interaction)
  - Task concordance (frequent tasks less steps)
  - Task visibility (show what users need)
  - Layout uniformity (moderately uniform and orderly layouts)
  - Visual coherence (related things put together)
Possible Design Metrics in Information Visualization

- Space usage efficiency
- Overlap extent
- Emphasized regions vs. other regions
- Color usage
- Connection between related views

Usability Inspection in UI

Figure 6.1: Survey with regard to frequencies of methods used and usefulness rating of them [Nie95]
We will look at

- Approach 1: Heuristic evaluation
- Approach 2: User testing

Heuristic Evaluation in UI

Definition:
- having a small set of evaluators to examine the interface and judge its compliance with recognized usability principles (the "heuristics").
Nielsen’s 10 Heuristics for UI (1)

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention

Nielsen’s 10 Heuristics (2)

- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation
How to Conduct a Heuristic Evaluation

- Heuristics
- Running prototype or simple screen dumps
- Three to five evaluators
- Evaluators work independently
- Output: A list of usability problems.
- Accumulate the results

Heuristic Evaluation

Advantages:
- Cheap
- Intuitive and easy to motivate people to do it
- Does not require advance planning
- Can be used early in the development process.
Formal User Testing VS. Heuristic Evaluation

- Expensive
- Discover local problems
- Usually must have running system
  - Formal
  - “Real users”, “Real tasks”
  - Performance metrics
- Cheap
- Discover global problems
- With or without running system
  - Informal
  - Can’t measure performance metrics

Formal User Testing (User Study)

Definition:
- A “large” set of “real” users performing “real” tasks in a working system
Essential Components of User Testing

- “Hard” components
  Working systems, environment, subjects, inspectors, datasets, tasks
- Motivation
  “There are ten possible approaches, I can only choose one”
  “Compare system A with system B”
- Hypothesis
  “A is better than B in C aspect”
- The testing methodology
- Information collection
  “average time users use to find D using A/B”
  Questionnaires and rating scales, video, screen capture
- Result analysis
  Statistic methods
  Visualization methods
  Multi-media analysis

Two Types of User Testing

Task-oriented user testing

- Users are asked to accomplish a number of tasks
  - “time out”
- Tasks: “Find Charlotte”, “Count 3s”, “Is A larger than B?”
- Results: speed, correctness…

Features:

- “Objective”
  - Easy to measure and compare
- “Rigid”
  - Task design is critical
Two Types of User Testing

Open-ended think aloud user testing
- Users are asked to find insights from data
- Insight: unit of discovery
  - Facts (actual finding), hypothesis, directed or unexpected…
- Results
  - distinct facts (actual findings), speed, domain value of insights, correctness
- Features:
  - Harder to analyze results
  - Broader

User Testing Step by Step

Step 1: Decision
- Motivation, hypothesis
- Figure out most important factors
  - A visualization system is complex!
  - Try to focus!
- Create a testing plan
User Testing Step by Step

Step 2: Preparation (It is time-consuming)
- Running prototype (minimize unrelated factors)
- Subjects (typical users)
- Hardware (room, projectors, camcorders, computers)
- Task sets or open-ended tasks (balanced)
- Subject information survey
- Training materials (presentation, training tasks)
- Questionnaires (rating scales)
- Open-end questions (feedback and clarification)

User Testing Step by Step

Step 2: Preparation – about result recording
- How to record results such as speed, correctness?
  - Automatically – accurate, less distractive for subjects
  - Manually
  - The more automatic, the better
    - Don't count on your own memory - you can't imagine how busy you will be in the experiment day!
    - Don't count on the users’ memories – they are as overwhelmed as you in the experiment!
    - You won't need to input many numbers from paper to computer if they are already there
User Testing Step by Step

Step 2: Preparation – pilot experiments
- Goal: calibrate the formal testing
  - Formal testing is expensive!
- Subjects: one or two users
- Process: iterative

User Testing Step by Step

Step 3: Experiment – general processes
- Approach 1:
  - Presentation (speaker)
  - Training (instructor)
  - Task performing (observer)
  - Questionnaire and open questions (listener)
- Approach 2:
  - Conductor sends experiment materials to subjects
  - Subjects learn by themselves
  - Subjects run testing by themselves
  - Subjects sent results to conductor
User Testing Step by Step

Step 3: Experiment – useful hints
- Alleviate subjects’ distressing (subject cried!)
- One-to-one vs. group testing.
- Don’t show individual subject’s name in report.
- Thank the subjects after the testing. Point out that they leads to improvement to your system.

User Testing Step by Step

Step 4: Result analysis and report
- Statistic analysis
- Visual exploration and representation
- Multi-media analysis
Example User Testing

- Motivation: Is HPC better than traditional PC?

- Methodology: use two groups of users, ask them to find as more insights as possible using HPC and PC respectively.
Example User Testing

- Preparing prototypes - HPC

Example User Testing

- Preparing prototypes - PC
Example User Testing

- Invite users, learn their backgrounds, dividing them into two balanced groups, and make time schedules (one day. Morning: PC. Afternoon: HPC)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>No. Subjects</th>
<th>CS-Majors</th>
<th>Prior Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPC</td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>FPC</td>
<td>11</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

- Reserve a large classroom. Reserve pizza!

Example User Testing

- Pilot experiments
  - They were very important!
Example User Testing

- Feedback Form
  - Name
  - OVERALL REACTIONS TO THE SOFTWARE
  - terrible wonderful
    - 0 1 2 3 4 5 6 7 8 9
  - difficult easy
    - 0 1 2 3 4 5 6 7 8 9
  - frustrating satisfying
    - 0 1 2 3 4 5 6 7 8 9
  - inadequate power adequate power
    - 0 1 2 3 4 5 6 7 8 9
  - dull stimulating
    - 0 1 2 3 4 5 6 7 8 9
  - rigid flexible
    - 0 1 2 3 4 5 6 7 8 9
- SCREEN
  - Characters on the computer screen
    - hard to read easy to read
    - 0 1 2 3 4 5 6 7 8 9
  - Highlighting on the screen simplifies task
    - not at all very much
    - 0 1 2 3 4 5 6 7 8 9
  - Organization of information on screen
    - confusing very clear
    - 0 1 2 3 4 5 6 7 8 9
- TERMINOLOGY AND SYSTEM INFORMATION
  - Use of terms throughout system
    - inconsistent consistent
      - 0 1 2 3 4 5 6 7 8 9

Open questions:
- If you have more time, do you think that you can still find more features?
- Is it hard or easy for you to find features using this system? Why?
- Do you think that there is anything we can do to improve this system?
- Are there any good points of this system?
- Any words you want to say to us...
Example User Testing

Result analysis
- Images saved with time stamps
  - Classify insights
  - Collect numbers and time stamps of insights
- Answers of open questions
- Answers of questionnaire

Figure 6.3: Pattern finding results of the HPC and FPC experiments [YWR03b].
Example User Testing

![Average Time Graph](image)

Figure 6.4: Average time interval in pattern finding in the HPC and FPC experiments [YWR03b].

Preferences