Hybrid Line-Based and Region-Based Interactive Set Data Visualization

Xiaohan Wang, Chuyu Zhang, Yu Zhu, Xueyi Chen, Liming Shen, Richen Liu*, and Rongtao Qian

Nanjing Normal University
Motivation

• Data exploration is popular:
  – It is challenging to visualize all set data items
• Existing methods are imperfect:
  – The **tabular** form
  – The **line-based** scheme
  – The **region-based** scheme
Background and Related Work

- Related Work
  - **Euler and Venn** diagrams
  - **Region-based** methods to visualize the set membership
  - **Line-based** methods to visualize the set membership

Simonetto et al.
Design Rationale

• Design Goals
  – Indicate the multiple categories which the set data item belongs to clearly.
  – Allow users to directly identify set data items belonging to an identical category.
  – Alleviates artefacts caused by empty overlapping regions in region-based methods without disconnected regions.
  – Avoid too much visual clutter while preserving the original layout of the timeline.
Algorithm 1: Balloon Connection

• The Connections Between Balloons

The connection process

The comparisons between the method using cost value
Algorithm 2: Line Connection

• Line Connection Algorithms
  – Consider each balloon as a rectangle
  – Add control points

```
function BA_Bloon_LINE_CONN(start_pt, end_pt, control_pts_list)
  while The line from start_pt to end_pt intersects the rectangle do
    Find the smaller area A
    if The top left or right point is in A then
      control_pts_list.add(the top left or right point)
      start_pt = the top left or right point
    else
      control_pts_list.add(the bottom left or right point)
      start_pt = the bottom left or right point
    end if
  end while
end function
```
Algorithm 3: Hierarchical Merge

- Hierarchical Clustering and Merging
  - Different depths of the tree indicate the different degrees
  - The number of nodes is the number of balloons automatically merged into this level.
User Study

• We recruited 12 participants
• The participants identify literatures cited in a survey paper by:
  – Line-based method
  – Region-based method
  – Our method

Age distribution

Post-study results
Case Study

• Conduct three evaluation tests on:
  – The survey of emerging trends of deep learning methods
  – The survey on simulation data visualization
  – The survey on social media data visualization

A survey paper on “emerging trends of deep learning methods”
Case Study

- Conduct three evaluation tests on:
  - The survey of emerging trends of deep learning methods
  - The survey on simulation data visualization
  - The survey on social media data visualization

A survey paper on “Simulation data visualization”
Case Study

• Conduct three evaluation tests on:
  – The survey of emerging trends of deep learning methods
  – The survey on simulation data visualization
  – The survey on social media data visualization

A survey on “Social media data visualization”
Case Study

- Case study on two cases:
  - Goddard Earth Observing System Model, Version 5 (GEOS-5)
  - Model of Ozone and Related Tracers, Version 4 (MOZART-4)
- Visualize **set information** in the overlapping clusters and overlapping simulation runs
Discussion and Future Work

• Limitations
  – Introduce visual clutter when the number of data items is too large
  – It is hard to perceive the colors when there are too many set categories

• Future Work
  – Enable users better control the merging processes interactively when hierarchical merging
Conclusions

• A novel method to explore set relations interactively in set data based on hybrid strategy: line-based and region-based
• Support placing the representative illustration for each set data item
• Reduce visual clutters
• The scalability is guaranteed
Acknowledgement

• The reviewers for their valuable comments
• All the user study participants
• Funding
  – National Nature Science Foundation of China (61702271)
  – Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX20_0445)
Thank You!