

Meredith Pulley (emailed sum and slides to Brad 11-01-06)  
10-16-06

**An insight-based methodology for evaluating bioinformatics visualizations**

Saraiya, P.; North, C.; Duca, K.;

Visualization and Computer Graphics, IEEE Transactions on  
Volume 11, Issue 4, July-Aug. 2005 Page(s):443 - 456

**Abstract:**

High-throughput experiments, such as gene expression microarrays in the life sciences, result in very large data sets. In response, a wide variety of visualization tools have been created to facilitate data analysis. A primary purpose of these tools is to provide biologically relevant insight into the data. Typically, visualizations are evaluated in controlled studies that measure user performance on predetermined tasks or using heuristics and expert reviews. To evaluate and rank bioinformatics visualizations based on real-world data analysis scenarios, we developed a more relevant evaluation method that focuses on data insight. This paper presents several characteristics of insight that enabled us to recognize and quantify it in open-ended user tests. Using these characteristics, we evaluated five microarray visualization tools on the amount and types of insight they provide and the time it takes to acquire it. The results of the study guide biologists in selecting a visualization tool based on the type of their microarray data, visualization designers on the key role of user interaction techniques, and evaluators on a new approach for evaluating the effectiveness of visualizations for providing insight. Though we used the method to analyze bioinformatics visualizations, it can be applied to other domains.

<http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1432690>

Overall, the group gave the article a good evaluation, except for a few issues with the details of the study's methodology. The group liked that the study introduced an "insight-based" methodology, but wanted an evaluation of which visualization (heat-map, parallel coordinates, cluster) was the best to understand or gain insight into each type of data set. In other words, which visualization or representation is most appropriate for a particular type of data? Instead of taking this approach, the study examined products (Spotfire, GeneSpring, HCE, TimeSearch, Clusterview). Too, the study lacked information/graphics explaining what each of these types of visualizations looked like. A follow-up of helpful links to accomplish this has been provided.

<http://array.mc.vanderbilt.edu/analysis/advanced.vmsr>

<http://cancerres.aacrjournals.org/cgi/content-nw/full/65/20/9226/FIG1>

[http://www.nature.com/nature/journal/v415/n6871/fig\\_tab/415530a\\_F1.html](http://www.nature.com/nature/journal/v415/n6871/fig_tab/415530a_F1.html)

<http://www.pnas.org/cgi/content/figsonly/102/2/261>

Furthermore, the study did not explain how the domain expert evaluated correct and incorrect insights. Perhaps a list of the expert generated insights for each data set could have been provided, as well as lists of user insights, to allow for comparison.