International Journal of Advanced Technology in Engineering and Science
Vol. No.3, Issue 11, November 2015
www.ijates.com
ijates
ISSN 2348 - 7550

A KNOWLEDGE FRAMEWORK FOR INTERACTIVE VISUAL ANALYSIS AND HIERARICHAL CLUSTERING TOOL

Harishbabu. Kalidasu[1], Battina Pullarao[2], A.Siles Balasingh[3]

[1, 2,3] Asst. Professor, Dept. of Computer Science

St.Jospeh University in Tanzania, Dar-es-salaam, Arusha Campus, Tanzania

ABSTRACT

Researchers understand about complex datasets to joining the visions from multiple coordinated visual displays which includes significant domain knowledge. Whenever deals with multidimensional data and clustering effects the most familiar displays and understandable remains one dimensional and 2-dimensional projections. Furthermore understand the displays of domain knowledge is smoother and ordered information for the same or correlated data sets. While unique parallel coordinates view[1] driven by a direct manipulation search which offers robust advantages but needs training for utmost researchers. Here in this paper we deliver a new tools and interaction examples about in what manner to incorporate users' domain knowledge for clear understanding of clustering results. The tools for visualization and analysis of cell pathways is to go outside limited static description of pathways besides to concern comprehensive networks which captures cellular dynamics. These tools effort to build pathways automatically and then add data from large datasets in proteomics, transcriptomics and genomics which contributes more significant demonstration of cell activity.

I. INTRODUCTION

The Structured databases, digital libraries and information spaces are the significant collections of modern-information environments. Searching a text is to locate a specific pages and initiating early points for exploration is extremely successful but this is not only first generation of knowledge discovery tools. Furthermore interfaces that maintain stability for datamining algorithms with effective information visualizations allow users to catch the significant clusters of appropriate documents, appropriate relationships between dimensions unusual outliers and amazing gaps[2]. The tools for visualization and analysis of biological pathways becomes more essential as the bottleneck in the research of cell biology which alters from data generating experimental step to data analysis and visualization step. At present the tools for cluster analysis are used for multidimensional data in various research areas which includes economical, financial, sociological, biological analyses.

Researchers in various areas are still evolving their own clustering algorithms although they are already a maximum number of general purpose clustering algorithms in presence. It is difficult to understand a clustering algorithm for their new data set. Furthermore most important reason stands it is difficult for researchers to authorize or recognize the clustering results in related to their knowledge of a data set. Eventhough the same clustering algorithm must generates absolutely different clustering results whenever the distance measure

changes. While the result could mark an intelligence to some of the researchers, but not to others because the legitimacy of a clustering result deeply depends upon the users' interest and its applications. So, researchers' domain knowledge plays a vital role while evaluating the particular clustering result.

In this paper we describe about some additions to interactive visual analysis tool i.e., Hierarchical Clustering Explorer (HCI) [3]. As these additions which include one dimensional histograms and 2-dimensional scatterplots accessing through coordinated views. These views are very known projections which are more logical than higher dimensional presentations. Hierarchical clustering explorer also produces performances of external domain knowledge.

As visualization techniques could be used to support some information extraction and semantic explanation for domain experts. For instance, visual analysis techniques such as active queries has been positively used in supporting researchers who is involved in analyses of multidimensional data, well-made visual co-ordination with researchers domain knowledge users accepting the result of particular analysis.

II. CLUSTERING

As we are having number of clustering algorithms such as k-means which requires users to state the number of clusters as an input, but it is tough to recognize the right number of natural clusters earlier. As other clustering algorithms will automatically determine how many number of clusters, then users might not be convinced the result, subsequently they required little or no control over the clustering process. Usually researchers have some information of each condition which based upon their expected clustering where related experimental conditions are clustered very closely by one another. A hierarichal clustering result is commonly characterized as binary tree called as "dendrograms" where subtrees are clusters. As this dendrograms can confirmed the outlook of related experimental cluster.

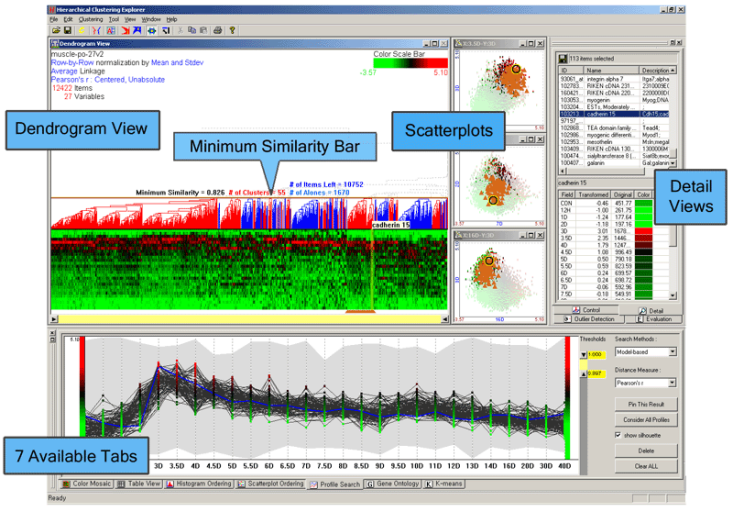

Fig 1. Source: Linked views in HCE (Seo 2006). (http://www.cs.umd.edu/hcil/hce/)

III. LINKING USERS DOMAIN KNOWLEDGE

Researchers like to group the genes with related profiles of gene expression or find stimulating time variant in the data set by accomplishing cluster analysis. As alternative way to identify the genes with their profiles are

related to well-known genes is to directly search for the particular genes by identifying the expected pattern of a well-known gene. While researchers consume domain knowledge such as predictable pattern of an earlier gene, here researchers can easily try to find out the genes which are similar to expected pattern meanwhile it is not easy to identify the expected pattern at single attempt, they need to conduct a series of searches for an gene expression profiles which are similar to expected pattern. They need interactive visual analysis tool which allows easily modifies an expected pattern and fast update of a search result.

Direct profile search and clustering can balance to each other because upto now there is no perfect clustering algorithm for all the datasets and applications. Direct profile search can be used to authorize the clustering outcome by projecting the search result against the clustering outcome observation. Usually, a clustering result could be used to authenticate the profile search by projecting the cluster outcome on the profile view. So, coordination between a direct search result and clustering result has variety of identification of process which are valid and effective.

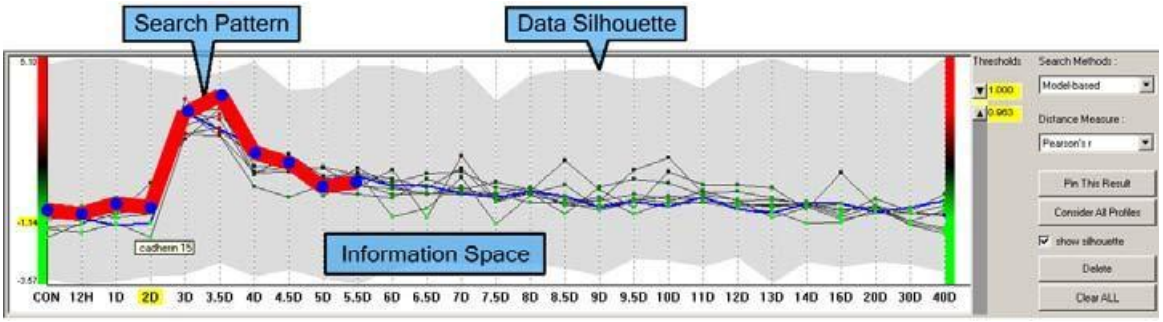

Fig 2. Source: HCE (Seo 2006) Allows Searching Gene Profile in Parallel Coordinates by Simply Drawing a Desired Gene Expression Profile. (http://www.cs.umd.edu/hcil/hce/)

Hierarichal clustering explorer repeats the functions of spotfires and time searchers with a different interfaces, the parallel coordinates view driven by a direct-manipulation search which allows a quick design and variation of a desired profiles by using different visual metaphors. As the time searcher supports collaborative querying and assessment of time-series data whereas the spotfire means profile search which calculates the comparison of search pattern for all the genes in a dataset and adds the outcome as a new column to the data set. As the fig 2 which consists of three parts of a parallel coordinates- the data space where input profiles are pinched and queries are stated, the range slider states similarity of thresholds as well as a set of controls states query parameters.

Genes included in the search outcome are stressed in the dendrograms view. On the other hand users check a cluster in the dendroram view, the profiles of the genes in the cluster will shown in the parallel coordinates view, therefore users can see the genes patterns in different view. For a large datasets, after an additional data points than pixels on the screen, a method that reduces heat map and dendrogram indication can be used. This methodology offers a way of simplifying the graphical presentation while maintain crucial information and provides maintenance for easy navigation and exhibit related information. This tactic relations can impress dendrogram and a detail-view dendrogram for each joined with a re-ordable heat-map. The overview exhibits only a user-controlled and partial number of nodes which represents the skeleton of a hierarchical manner.

As the tools in the group provides the capability to discover relationships, gaps, outliers, clusters and other futures in the data. Though various methods were adopted from a non-biological areas, it is a major challenge

for accomplishment to biological meaning from exposed relationships. As one of the researcher saraiya gives some drawbacks about clustering approaches which can be potentially preferred the users into a particular line of held too quickly.

IV. CONCLUSION

This paper presents about two co-ordinates views to integrate users' domain knowledge through visual analysis of the dataset and clustering results. While users known an estimated pattern of a different nominee group of interest and they can use parallel co-ordinates view rapidly unite the search pattern according to their domain knowledge besides run a direct manipulation search. The efforts that are designed to help users performs empirical data analysis form a meaningful assumptions and verify results. The visualization methods can help molecular biologists analyze their multidimensional gene expression profile data.

While the visualization and analysis tools are biological pathways which effects towards bigger and richer pathway visualizations integrate their high throughput experimental data and information from various databases of a gene expression.

REFERENCES

[1]. Classification and visualization for high-dimensional data by Inselberg, A., Avidian, 6th ACM International Conference on Knowledge discovery and data mining (2000) pg.370-374.

[2]. Combining information visualization with datamining, proc. discovery science 4th international conference 2001, Shneiderman. B

[3]. Interactively exploring hierarchical clustering results by Seo. J. Shneiderman, IEEE computer, vol. 35, no.7 (2002) pg.80-86.

[4]. Dynamic query tools for time series data sets by Hochhesier, H., Shneiderman, B. vol 3. Pg 1-18

[5]. Shneiderman, B., Inventing discovery tools: Combining Information Visualization with Data Mining, Proc. Discovery Science 4th International Conference 2001, Editors (Jantke, K. P. and Shinohara, A.), Springer-Verlag, Berlin Heidelberg New York (2002) 17-28. Also printed in Information Visualization, Vol. 1. (2002) 5-12

[6]. Saraiya, P. et al. "Visualizing biological pathways: requirements analysis, system evaluation and research agenda." Information Visualization 4, no. 3 (2005): 191-205.

[7]. Zhou, H. et al. "Visual Clustering in Parallel Coordinates." Computer Graphics Forum 27, no. 3 (2008): 1047 - 1054.