Multiple Foci Drill-Down through Tuple and Attribute Aggregation Polyarchies in Tabular Data

Nathan Conklin, Sandeep Prabhakar, Chris North
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
{nconklin, sprabhak, north}@cs.vt.edu
http://infovis.cs.vt.edu/

Abstract

Information analysis often involves decomposing data into sub-groups to allow for comparison and identification of relationships. Breakdown Visualization provides a mechanism to support this analysis through user guided drill-down of polyarchical metadata. This metadata describes multiple hierarchical structures for organizing tuple aggregations and table attributes. This structure is seen in financial data, organizational structures, sport statistics, and other domains. A spreadsheet format enables comparison of visualizations at any level of the hierarchy. Breakdown Visualization allows users to drill-down a single hierarchy then pivot into another hierarchy within the same view. It utilizes a fix and move technique that allows users to select multiple foci for drill-down.

1. Introduction

In many domains, data analysts work to identify patterns within large data sets and then attempt to discover events that may have caused the pattern to occur. For example, financial analysts attempt to identify key events that have enabled a particular company to perform well. Similarly, sports analysts attempt to identify causes for a team's success. Meanwhile, the competition continues to look for weaknesses. Both financial and sports analysts need to be able to drill-down through the data in an attempt to identify behavior causalities.

Data aggregation is typically required to generate an overview of very large data spaces [12]. The aggregation provides a reduced data representation, allowing analysts to identify patterns. They need to breakdown this overview. An *aggregation hierarchy* divides data from a superset into subsets. The parent node is the superset (aggregated data) and the children nodes are the subsets (subgroup details). Analysts *drill-down* from aggregated overviews to details. By comparing the detailed subsets to the overview, analysts can understand what factors affect the overall behavior.

Most data sets can have multiple aggregation hierarchies within them. For example, the financial market can be decomposed hierarchically based on sector, geography,

capitalization, product types, or financials. Each company is a leaf node in each of these hierarchies. These multiple aggregation hierarchies, intersecting at the leaf nodes, define a financial market *aggregation polyarchy* (a special case of polyarchy [9]).

We define a new problem, called *tabular polyarchies*, in which tabular data contains aggregation polyarchies on both its tuples and its attributes (see Figure 1):

- *Tuple polyarchy*: tuples are aggregated by grouping the data based on categorical information.
- *Attribute polyarchy*: attributes are aggregated using mathematical formulae similar to spreadsheets.
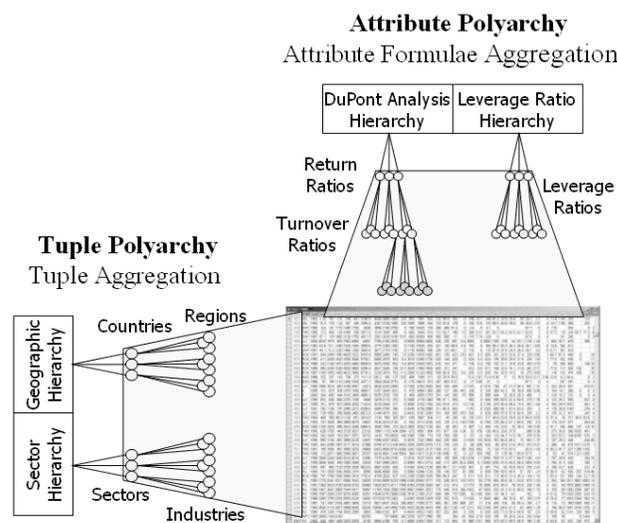

Figure 1: Polyarchical structure on the company tuples and ratio attributes of tabular financial data.

Furthermore, each data value typically is a time series, adding a third dimension to the table. This organization is common in many application domains. In the financial analysis example, an analyst has the task of organizing the data in an attempt to identify the behaviors that affect company performance. This technique for financial statement analysis is called DuPont Analysis [10]. Typically, the DuPont Analysis ratio tree (Figure 2) is used to lead the analysis.
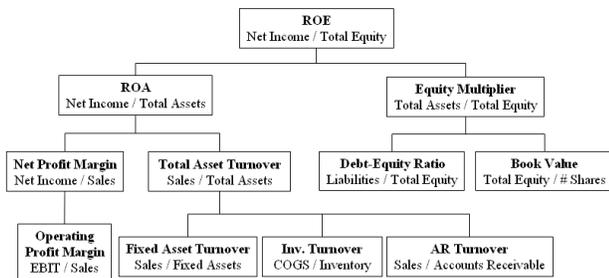
Figure 2: DuPont ratio formula hierarchy used to aggregate financial attributes.

The analyst starts by looking at the Return on Equity (ROE) values for a company. They compare the value over time to understand the direction the company is headed. To understand which causalities affect the behavior of the ROE, the user needs to drill-down through the ratio hierarchy. By comparing children ratios to the parent ratio, the analyst can identify which financial indicators have the largest affect on the parent's behavior. It is important to identify the lowest level causalities because they can drive management decisions and direct the company's planning.

2. Related Work

Hierarchies are common information structures used when organizing related items. Many visualization techniques have evolved for visualizing and interacting with hierarchies. Recent research examines multiple related hierarchies. Multi-trees [5], Time Tube [3], and Taxonomy Visualization [6] each examine a different class of multiple related hierarchies. Polyarchy Visualization [9] introduces a *general polyarchy* as multiple intersecting hierarchies that share at least one node. The visualization introduces the idea of a *visual pivot*. From any node in a hierarchy, the user can pivot to that node's position in an intersecting hierarchy. This allows the user to understand the multiple hierarchical contexts of selected nodes.

TableLens [8], FOCUS [13], and others demonstrate the natural advantages of the visual tabular format for visualizing tabular data. Visualization Spreadsheet [2] extends the tabular metaphor by expanding each cell into a miniature visualization of a dataset. This enables the visual comparison and manipulation of related datasets.

Some tools support visualization of aggregated tuples. Visage [11] demonstrates the use of drill-down through categorically aggregated tuples. Similarly, Polaris [14] utilizes a PivotTable [7] approach, and various OLAP tools provide similar support.

Breakdown Visualization extends this work by supporting drill-down through both tuple and attribute formulae aggregation. Breakdown Visualization combines polyarchical drill-down with a spreadsheet format to visually organize the data. The company polyarchy is generated from the Standard & Poor's classifications that are provided as part of the COMPUSTAT database [4]. The ratio polyarchy utilizes ratios identified in DuPont Analysis [10].

3. Breakdown Visualization

Breakdown Visualization allows for both column-based formulae aggregation and row-based tuple aggregation. It utilizes a polyarchical drill-down path to allow for deep drill-down in either dimension. Breakdown Visualization supports DuPont Analysis by enabling users to navigate through the metadata polyarchies while the details are displayed in a spreadsheet format (Figure 3). This format allows the user to identify causal relationships by comparing the overview to the details. Similarly, users can evaluate alternatives by comparing multiple detail views.

3.1. Layout and Polyarchy Drill-Down

Breakdown Visualization uses the indented hierarchical format to show the drill-down path. This hierarchy only shows the drill-down path and bottom-level leaf nodes. Nodes from the attribute polyarchy act as column headings. They identify the attribute formulae that determine the financial metric displayed in the column. The formulae shown in Figure 2 are related to each other because the parent ratio is derivable from its children ratios. Similarly, the company polyarchy nodes act as row headings. They identify the tuple-based groupings that determine the values for each row. As a user drills down through the hierarchy, they break down the aggregated datasets into their subgroups. Heading color helps to reveal hierarchy levels.

The spreadsheet layout allows for comparisons between companies and ratios. Each cell provides a time series visualization of the ratio value for the company over time. Visualizations for aggregated tuples are generated by visually overlaying the time series of each aggregated company. Users can choose either a scatter plot, line graph, or histogram. A TableLens format could be used to visualize data that does not have the time series dimension.

Breakdown Visualization allows the user to drill-down on both dimensions of the table. The user selects the path to traverse by clicking on the row or column heading and selecting a drill-down path from a pop-up menu. Figure 3 demonstrates the pop-up menu and choice of drill-down paths. As the user traverses deeper into the hierarchy, the sibling nodes on earlier levels disappear to support scale up. The drill-down path always remains in view.

Breakdown Visualization allows drill-down across hierarchies by allowing users to *pivot* into other hierarchies. This pivot is similar to the PivotTable, breaking down the current level based on the child nodes defined in the new hierarchy (Figure 4). However, this

pivot is different than the visual pivot introduced by Robertson, et. al. [9]. In Breakdown Visualization, the new children are shown as the next level in the current drill-down path.
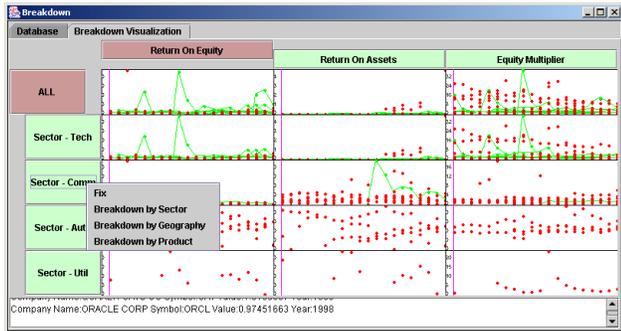

Figure 3: Breakdown Visualization showing an initial drill-down in the tuple and attribute polyarchies.

Figure 4 shows the pivot from sector to geography within the company polyarchy. The ability to pivot into separate hierarchies allows the analyst to relate different aspects of the data. For example, a financial analyst may traverse a path similar to Figure 4 to determine if the recent recession that affected the communications sector was centralized within a specific geographic region.
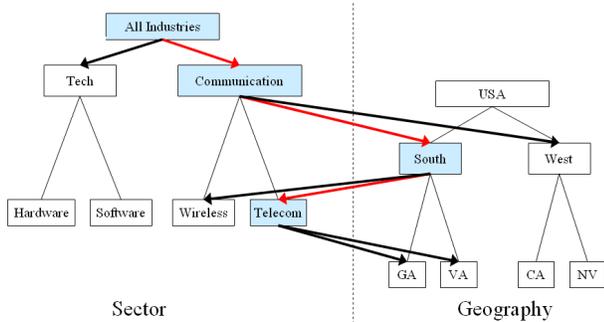

Figure 4: Pivoting across aggregation hierarchies allows the user to drill-down based on different categorical values.

Support for coordinated brushing across the separate views allows the user to identify and relate the patterns between different cells [1]. As shown in Figure 3, when a user selects multiple points colored lines are drawn that connect the time series data allowing the user to identify changes over time. Corresponding parent and children rows highlight the same data points. This allows the user to compare the tuple aggregated overviews with the drilled down detail views. As points are selected, further details are shown on demand in the detail window at the bottom of the visualization.

3.2. Multiple Foci Drill-Down Paths

It is common for analysts to need multiple focus points within their drill-down path. Pattern identification may indicate multiple children nodes showing a similar behavior. The analyst needs to drill-down and compare

each of these paths in order to determine the events that are driving the behavior.

The *fix and move technique* allows the users to establish multiple foci. The user selects a node to fix in either the company or ratio polyarchy. They then drill down on a different node while the drill-down hierarchy retains the fixed node. He/she is then able to start a separate drill-down path by later selecting the fixed node for breakdown as shown in Figure 5. This technique adds the capability of multiple foci within the same drill-down hierarchy.
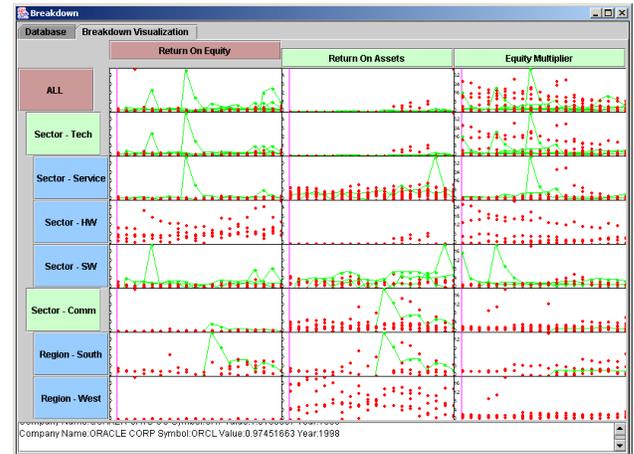

Figure 5: Multiple foci drill-down in the tuple polyarchy enables comparison across different drill-down paths.

For each focal point, Breakdown Visualization maintains the last drill-down location in each aggregation hierarchy. This is demonstrated in Figure 4, where pivoting from the 'South' region in the geography hierarchy back into the industry hierarchy resumes the drill-down by breaking down the 'Communications' sector. This is important because the current path previously drilled-down through 'Communications', filtering all other sectors, so starting back at 'All Industries' would not make sense.

Figure 6: Drill-down in the attribute polyarchy allows an analyst to identify factors affecting overall ROE.

Users can interactively explore the data through drill-down operations in each polyarchy. They can examine several drill-down paths by fixing and unfixing nodes. This allows them to discover deeper relationships by splitting the analysis in multiple directions.

4. Limitations and Future Work

Continued work is in progress to improve the capabilities of the user interface. This includes allowing the user to define the aggregation sequences that construct the tuple aggregation polyarchy similar to the PivotTable. Support is also needed to allow users to define customized formulae during analysis. This may include the ability to define formula by combining data primitives.

Better visualization of the polyarchical structure is needed. Integrating the polyarchical structure along with the drill-down path may provide users with better awareness when analyzing data. Breakdown Visualization scales up to a few hundred cells on desktop displays. However, trees with a high branching factor generate a large number of cells and can make it difficult to interpret the time series visualizations.

The polyarchical drill-down is a new approach for data exploration. Further studies are needed to evaluate its capabilities and differences when compared to visual pivot tables.

5. Conclusion

Breakdown Visualization introduces the following contributions:

- Definition of the tabular polyarchies problem that combines categorical tuple aggregation polyarchy and attribute formulae aggregation polyarchy to organize tabular data, and integrates a time series dimension.

- A user interface that integrates a spreadsheet layout with polyarchy navigation of rows and columns. Support for drill-down from overview to details along with pivoting across hierarchies. A fix and move technique allowing for multiple foci.

Breakdown Visualization allows users to compare the parent and children aggregates for both tuples and attribute formulae. This provides users with the ability to understand causal relationships that affect overall behavior.

Breakdown Visualization can also be applied to tabular polyarchy data that does not have the time-series dimension by replacing the visualization spreadsheet with a TableLens. It can be applied to several application domains including financial analysis, organizational structures, sports statistics, and other fields.

References

1. Becker, R. and Cleveland, W. Brushing scatterplots. *Technometrics*, 29(2), pp. 127-142, 1987.

2. Chi, E., Barry, P., Riedl, J., Konstan, J., A Spreadsheet Approach to Information Visualization. *Proc. IEEE Information Visualization*, pp. 17-24, 1997.

3. Chi, E., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R., and Card, S. Visualizing the evolution of web ecologies. *Proc. ACM SIGCHI*, pp. 400-407, 1998.

4. COMPUSTAT Industrial Annual database, http://www.compustat.com/, *Standard and Poor*, 2001.

5. Furnas, G., and Zacks, J., Multitrees: Enriching and reusing hierarchical structure. *Proc. ACM SIGCHI*, pp. 330-336, 1994.

6. Graham, M., Kennedy, J., and Hand, C., A comparison of set-based and graph-based visualizations of overlapping classification hierarchies. *Proc. ACM AVI*, pp. 41-50, 2000.

7. Microsoft Excel – User's Guide, Microsoft, Redmond, WA, 1995.

8. Rao, R., and Card, S. The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. *Proc ACM SIGCHI*, pp. 318-322, 1994.

9. Robertson, G., Cameron, K., Czerwinski, C., Robbins, D., Polyarchy Visualization: Visualizing Multiple Intersecting Hierarchies. *Proc. ACM SIGCHI*, 2002.

10. Ross, S., *Fundamentals of Corporate Finance*, McGraw Hill, 2000.

11. Roth, S., Lucas, P., Senn, J., Gomberg, C., Burks, M., Stroffolino, P., Kolojejchick, J., Dunmire, C., Visage: A user interface environment for exploring information. *Proc IEEE Information Visualization*, pp. 3-13, 1996.

12. Spence, R., *Information Visualization*, ACM Press, 2000.

13. Spenke, M., Beilken, C., and Berlage, T. FOCUS: The Interactive Table for Product Comparison and Selection. *Proc ACM UIST*, 1996.

14. Stolte, C., and Hanrahan, P., Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. *Proc IEEE Information Visualization*, 2000.