Visual Data Mining with ILOG Discovery

Thomas Baudel, Bruno Haible, and Georg Sander

ILOG SA, 9 rue de Verdun - BP 85, 94253 Gentilly Cedex, FRANCE

{baudel|haible|sander}@ilog.fr

<http://www2.ilog.com/preview/Discovery>

1 Introduction

Data mining deals with the discovery of useful and previously unknown knowledge from large data sets [3]. Traditional data mining tools use a combination of machine learning, statistical analysis, modeling techniques and database technology to find patterns, exceptions and subtle relationships in data. Typical applications include market segmentation, customer profiling, fraud detection, credit risk analysis, and business data development.

ILOG Discovery is different to traditional data mining tools: It is an interactive visual information exploration tool. From data sets gathered from your information systems (such as spreadsheets, statistical or usage databases, usage logs or accounting data), ILOG Discovery provides simple yet powerful ways to view this data in many different ways. It allows you to quickly change the display parameters, so that you can interactively search for views of your data that highlight trends, spot general patterns, or pinpoint exceptions. ILOG Discovery can easily handle large amounts of data and produce a wide variety of visualizations with very few manipulations, which makes it a completely novel exploration tool.

2 Highlights

ILOG Discovery reads a set of records, each record containing many fields. It displays the records while mapping the fields to different visual parameters such as color, width, position, etc. The mapping methods include:

- 2-D Charts: traditional x-y charts that allow to quickly exchange and adapt the x and y axes and visual parameters (color, size, chart layout).
- Parallel histograms (similar to table lens [4]) that allow to sort, filter and compare different fields of the records.
- Clustered Treemaps [2,5] where the nested clustering is shown by an inclusion drawing standard while size, color, position and structure of each cluster can be used to express certain aspects of the data.
- and many combinations of these techniques.

While ILOG Discovery's technical origin is more close to 2-D charting, it uses some graph layout techniques to manage clustered treemaps. A treemap

Fig. 1. ILOG Discovery

consists of clusters that are shown as rectangular areas. The cluster nesting is an inclusion tree that must be laid out with respect to space constraints, e.g. each cluster should approximately cover a square, and the area for the entire drawing is given, similar to [1]. Also, ILOG Discovery uses a simple still powerful labeling technology to annotate the diagrams.

ILOG Discovery is implemented in pure Java. A stand alone preview version can be downloaded for free at <http://www2.ilog.com/preview/Discovery>.

References

1. P. Eades, Lin T., and Lin X. Two tree drawing conventions. *Int. Journal of Computational Geometry and Applications*, 3(2):133–153, 1993.
2. J.-D. Fekete and C. Plaisant. Interactive information visualization of a million items. In *Proc. IEEE Symp. InfoVis 2002*, pages 117–124. IEEE Press, 2002.
3. D.J. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. MIT Press, 2001.
4. R. Rao and S. K. Card. The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proc. ACM SIGCHI'94*, pages 318–322. ACM Press, 1994.
5. B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99, 1992.