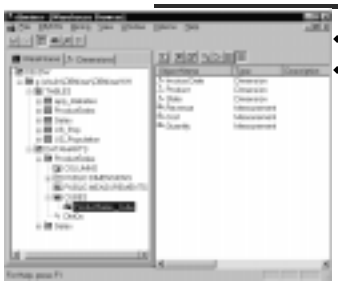
# Evaluation of DBMiner

By:
Shu LIN
Calin ANTON

## Outline

- Importing and managing data source
- Data mining modules
  - Summarizer
  - Associator
  - Classifier
  - Predictor
- Conclusions

## View of Warehouses and Hierarchies



- Tables
- ***Datamarts***
  - Columns
  - Dimensions
  - Measurements
  - ***Cubes***

## Creating Data Warehouse

- Importing relational data: only the default database that DBMiner provides can be used.
- Easy to import a datamart and difficult to create one.
  - Datamart can be associated with only one table/view.
  - Creation of datamart is not easy enough for usual users and not flexible enough for professional users

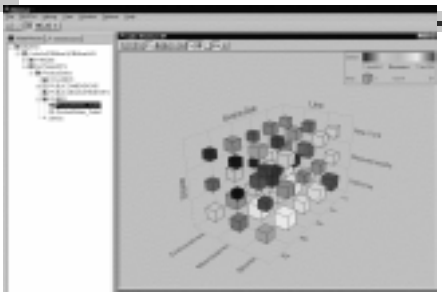
## Creating Data Warehouse (continued)

- Concept hierarchy is automatically generated when a dimension is created, but the user can scarcely customize it.
- Only numerical attributes are allowed as measures
- We think it would be better if DBMiner had a wizard to help user build a data cube.

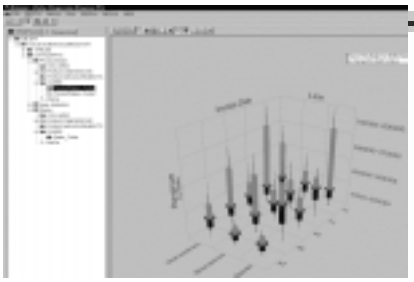
## Browsing Data Cube

- The visualization of data cube is very instructive.
- User can perform some OLAP operations but slicing.
- Data cube browser also presents the data dispersion.

## Browsing Data Cube (continued)



## Browsing Data Cube (continued)



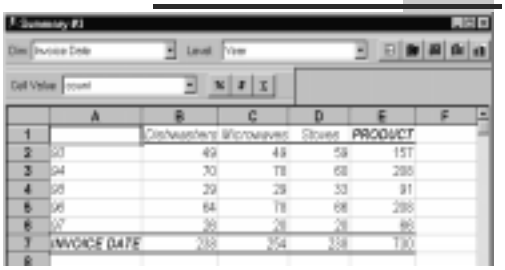
## Data Mining

- Mining functions: summarization, association, classification, prediction.
- OLAP functions are integrated with mining functions.
- The mining process is performed on line.
- A wizard guides user through mining process.

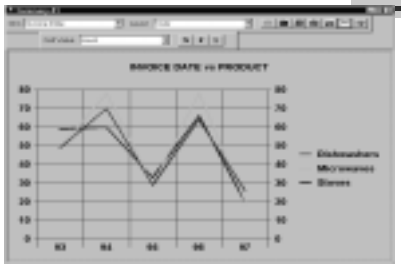
## The Summarizer

- Generalize/summarize data at high abstraction levels
- The output is presented in six different forms: crosstab, 3D bar chart, 3D area chart, 3D cluster bar, 2D bar chart and 2D line char.
- The output can be presented on only two dimensions at a time, and not all the combinations of two dimensions are possible.

## The Summarizer (continued)



## The Summarizer (continued)

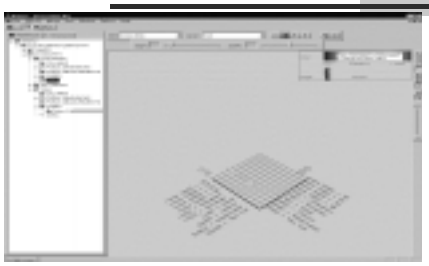


## The Associator

- Find association among a set of attributes and their values.
- We find a bug in this module.
- User can change the settings or constraints during association mining to make the association rules more accurate.
- Association rule is visualized as table, bar chart, and ball chart.

## The Associator (continued)

## The Associator (continued)



## The Associator (continued)

## The Classifier

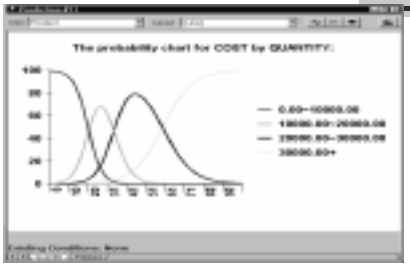
- Based on the features present in the class_labeled training data, develop a description or model for each class.
- The output is presented as a complete classification tree (decision tree)
  - good in the sense that user can get a clear impression of classification process.
  - Redundant if user only cares about the classification rules.

## The Classifier (continued)

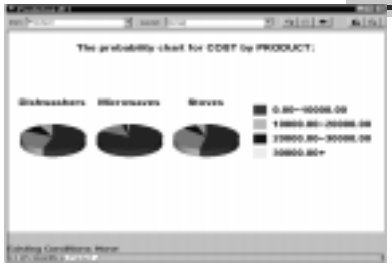
## The Predictor

- Predict data value distributions based on the available data.
- The output is presented as a set of curves if the predictive attribute has continuous numeric values; otherwise, a set of pie charts is used.
- The output is present only on one dimension. It cannot show how the combination of two predictive attributes affect the predicted attribute.

## The Predictor (continued)




## The Predictor (continued)




## Conclusions

- It integrates OLAP functions with mining functions.
- It works on line, i.e., it is fast.
- It generates multiple forms of output: graphics, tables, and different kinds of charts;
- It has a user-friendly interface, and for each mining function it has a wizard, which guides the user through the mining process.