SNS COLLEGE OF ENGINEERING
Coimbatore-107

UNIT III

DATA MINING

DATA MINING TASK PRIMITIVES

By

T.R.Lekhaa

AP-IT

SNSCE

Introduction

- Can specify DM task in the form of DM query

- These primitives allows to communicate in an interactive manner with DM system

- List of DM Task Primitives:
  - Set of task relevant data to be mined
  - Kind of knowledge to be mined
  - Background knowledge
  - Interestingness measures and thresholds for pattern evaluation
  - Representation for visualizing the discovered patterns

Primitives for specifying a data mining task

- **Task-relevant data:** This primitive specifies the data upon which mining is to be performed.

- It involves specifying

  - the database and tables or data warehouse containing the relevant data, conditions for selecting the relevant data, the relevant attributes or dimensions for exploration, and instructions regarding the ordering or grouping of the data retrieved.

- **Knowledge type to be mined:** This primitive specifies the specific data mining function to be performed, such as characterization, discrimination, association, classification, clustering, or evolution analysis.

- As well, the user can be more specific and provide pattern templates that all discovered patterns must match. These templates or meta patterns (also called meta rules or meta queries), can be used to guide the discovery process.

- **Background knowledge:** This primitive allows users to specify knowledge they have about the domain to be mined.

- Such knowledge can be used to guide the knowledge discovery process and evaluate the patterns that are found.

- The several kinds of background knowledge, this chapter focuses on concept hierarchies.

- **Pattern interestingness measure:** This primitive allows users to specify functions that are used to separate uninteresting patterns from knowledge and may be used to guide the mining process, as well as to evaluate the discovered patterns.

- This allows the user to confine the number of uninteresting patterns returned by the process, as a data mining process may generate a large number of patterns.

- Interestingness measures can be specified for such pattern characteristics as simplicity, certainty, utility and novelty.

- **Visualization of discovered patterns:** This primitive refers to the form in which discovered patterns are to be displayed.

- In order for data mining to be effective in conveying knowledge to users, data mining systems should be able to display the discovered patterns in multiple forms such as rules, tables, cross tabs (cross-tabulations), pie or bar charts, decision trees, cubes or other visual representations.