**7.** The **scope** of frequent pattern mining research reaches far beyond the basic concepts and methods introduced in Chapter 6 for mining frequent item-sets and associations. This chapter presented a road map of the field, where topics are organized with respect to the kinds of patterns and rules that can be mined, mining methods, and applications.

In addition to mining for basic frequent item-sets and associations, **advanced forms of patterns** can be mined such as multilevel associations and multidimensional associations,quantitative association rules, rare patterns, and negative patterns. We canalso mine high-dimensional patterns and compressed or approximate patterns.

**Multilevel associations** involve data at more than one abstraction level (e.g., “buys computer” and “buys laptop”). These may be mined using multiple minimum support thresholds.

**Multidimensional associations** contain more than one dimension.

Techniques for mining such associations differ in how they handle repetitive predicates.

**Quantitative association rules** involve quantitative attributes. Discretization, clustering, and statistical analysis that discloses exceptional behavior can be integrated with the pattern mining process.

**Rare patterns** occur rarely but are of special interest. **Negative patterns** are patterns with components that exhibit negatively correlated behavior. Care should be taken in the definition of negative patterns, with consideration of the null-invariance property. Rare and negative patterns may highlight exceptional behavior in the data, which is likely of interest.

**Constraint-based mining** strategies can be used to help direct the mining process toward patterns that match users’ intuition or satisfy certain constraints. Many users specified constraints can be pushed deep into the mining process. Constraints can be categorized into **pattern-pruning** and **data-pruning** constraints. Properties of such constraints include monotonicity, antimonotonicity, data-antimonotonicity, and succinctness.

Constraints with such properties can be properly incorporated into efficient pattern mining processes.

Methods have been developed for mining patterns in **high-dimensional space**. This includes a pattern growth approach based on row enumeration for mining data sets where the number of dimensions is large and the number of data tuples is small (e.g., for microarray data), as well as mining **colossal patterns** (i.e., patterns of very long length) by a Pattern-Fusion method.

To reduce the number of patterns returned in mining, we can instead mine compressed patterns or approximate patterns. Compressed patterns can be mined with representative patterns defined based on the concept of clustering, and approximate patterns can be mined by extracting **redundancy-aware top-k patterns** (i.e., a small set of k-representative patterns that have not only high significance but also low

Redundancy with respect to one another).

**Semantic annotations** can be generated to help users understand the meaning of the frequent patterns found, such as for textual terms like “ffrequent, patterng.” These are dictionary-like annotations, providing semantic information relating to the term.

This information consists of context indicators (e.g., terms indicating the context of that pattern), the most representative data transactions (e.g., fragments or sentences containing the term), and the most semantically similar patterns (e.g., “fmaximal, patterng” is semantically similar to “ffrequent, patterng”). The annotations provide a view of the pattern’s context fromdifferent angles, which aids in their understanding.

Frequent pattern mining has many diverse applications, ranging from pattern-based data cleaning to pattern-based classification, clustering, and outlier or exception analysis.

These methods are discussed in the subsequent chapters in this book.