DOMAIN-CONCEPT MINING:
AN EFFICIENT ON-DEMAND DATA MINING APPROACH

Wannapa Kay Mahamaneerat
Dr. Chi-Ren Shyu, Dissertation Supervisor

ABSTRACT

Traditional brute-force association mining approaches, when applied to large datasets, are thorough but inefficient due to computational complexity. A low global minimum probability threshold can worsen this complexity by producing an overwhelming number of associations; however, a high threshold may not uncover valuable associations, especially from under-represented groups within the population. Regardless, the uncovered associations are not systematically organized.

To solve these problems, novel Domain-Concept Mining (DCM) with Partition Aggregation (DCM-PA) has been developed. DCM organizes data by grouping transactions with common characteristics, such as a certain age group, into “domain-concepts” (dc). DCM granulizes partitioning criteria by pairing each attribute with its values. Criteria may include under-represented groups as well as spatial, temporal, and incremental dimensionalities. Then, a statistical power analysis is utilized to determine if multiple criteria of the same attribute, such as age group 18-24 and 25-34, should be combined to form a broader partition. Doing so maintains the tradeoff between findings with statistical significance and computational resource consumptions, while preserving data organization. Associations can be extracted from each partition independently because a partition contains all of its qualified transactions. Moreover, the partition size proportionally adjusts the global threshold to be more specific and sensitive.

After the initial phase is complete, DCM-PA efficiently reuses DCM’s associations to compute results from multiple-partition aggregation (union or intersection) using Bayes Theorem and a pipelining technique. DCM-PA offers the flexibility to perform association mining that is expected to uncovering more valuable knowledge through means like trends and comparisons from various dc partitions and their aggregations.