
Abstract submission for invited talk: David Chiu, University of Guelph

Title: Multi-value Association Patterns and Their Applications to Bioinformatics

Abstract:

Identifying significant relationships even in data when multiple sources are involved is extremely important in bioinformatics and systems biology. For example, linking genotype and phenotype attributes plays a vital role in understanding biological processes, including cellular and organismal responses under normal or novel conditions such as in diseases or genetic modifications. At the level of biomolecules, identifying relationships between certain sequence segments or molecular sites and their effects on structure and biological functionalities is a very active area of research. To gain more insights into the internal interdependency relationships between data values, different types of high-order (multiple values) association patterns have been proposed. Recently, special types we called Nested High-Order Pattern (or NHOP) and Consigned Pattern (CP) are evaluated. The general form of High-Order Pattern (HOP) is a set of multiple associated values (identified as selected variable outcomes) extracted from a random $N$-tuple. The pattern generalizes sequential pattern with contiguous values and reflects more specific relationships than multiple-variable pattern. The pattern is detected by statistical testing if the occurrence is significantly deviated from the expected according to a prior model or null hypothesis. We have evaluated our work for multiple aligned sequence analysis as well as to classification tasks. The motivation is that, association patterns that can be extracted using a relevant significance criterion and integrated into different tasks can reinforce the reliability, and hence provide a better understanding of the data domain. In this talk, we will discuss different experiments involving biomolecules, including the P53 cancer gene and a protein family known as SH3 domain, a model molecule for protein interactions.