Time series data identification and optimization with the measurements of potential bias and discrimination issues based on pseudo-Boolean functions

Jinwook Lee, Lanqing Du, Jade Zhang, Matthew Schneider
Decision Sciences and MIS, Drexel University, Philadelphia, PA 19104, United States

Abstract

Suppose that we are given N time series data X_i for $i = 1, \ldots, N$, where each time series is assumed to be m-dimensional and weak stationary. Let $A = \{X_1, \ldots, X_N\}$ be the set of such given time series data. Then A is a set of m-vectors and we typically have $m \ll N$. The set A can be represented, using m observations from each of the N time series, as an $m \times N$ data matrix X (i.e., N columns of m-vectors). Based on the data matrix, we consider given time series vectors as nodes and their pairwise relationships as edge weights. Then we can draw a graph $G = (V, E)$, which is a complete graph K_N, where we have N nodes and $N(N-1)/2$ edges. Using the graph, we formulate novel optimization problems based on suitable weight functions for aggregation and identification of target features. This paper contributes mainly to three literatures: the growing body of work on data with potential discrimination and bias issues; the study of data aggregation; and their counterparts on the graph. There are a wide range of metrics to measure unfairness and discrimination among different groups. We use pseudo-Boolean functions for the generalization of such metrics, that are used for effective detection and aggregation of data. Numerical examples are presented.