An Integer Programming formulation for the Polymorphic Alu Insertion Recognition Problem

Luciano PORRETTA
Université libre de Bruxelles

A point-interval \((I_v, p_v)\) is pair constituted by an interval \(I_v\) of \(\mathbb{R}\) and a point \(p_v\) \(\in I_v\). A graph \(G = (V, E)\) is a Max-Point-Tolerance (MPT) graph if each vertex \(v \in V\) can be mapped to a point-interval in such a way that \((u, v)\) is an edge of \(G\) iff \(I_u \cap I_v \supseteq \{p_u, p_v\}\). MPT graphs constitute a superclass of the interval graphs and naturally arise in genome-wide association studies as a way to represent relationships of loss of heterozygosity among DNA fragments extracted from a population of individuals. In this article we investigate the minimum cost clique cover problem in a MPT graph. This problem is also known in the literature as the Parsimonious Loss of Heterozygosity Problem (PLOHP) and is proved to be \(\mathcal{NP}\)-hard. The optimal solution to the PLOHP has a remarkable practical importance as it enables the association of major human diseases with chromosomic regions from patients that underwent to loss of heterozygosity events. We present a integer linear programming formulation for the PLOHP based on column generation and we develop a number of preprocessing techniques and decomposition strategies to reduce the size of its instances. Computational experiments show that the proposed algorithm is 10-30x faster than the previous approaches described in the literature, and suggest, at the same time, new directions for the development of future exact solution approaches that may prove of fundamental assistance in practice.

March 1, 2016 – 4:30 p.m. – Room b - 135