

Profitable product set discovery

Consider a bipartite graph $G(U,V,E)$ modeling the following real-world problem. The set U represents a set of products for sale. The set V represents the set of all previously purchased bundles of products. The set E represents the sets of pairs (u,v) denoting that product u from U was sold in a bundle v from V . Furthermore, for each bundle v , we know the profit $w(v)$ associated to v (e.g., the sum of the profit of each product in v). Now consider that we would like to select a sub-collection of bundles that have a largest ratio between the total profit and the number of distinct products in them. We call this the profitable product set discovery (PPSD) problem. Intuitively, we would like to select a set of heavy nodes in G that live in a small neighborhood. Discovering such bundles can be greatly beneficial to a retailer, as they often wish to know the set of products that bring a large profit when sold in specific bundles. This is because a retailer can buy such products cheaply in bulk, or transport and store them with lower costs. In this project, we plan to design exact or approximate algorithms to solve PPSD. We also plan to implement these algorithms and test them using real benchmark datasets. To this end, we are looking for someone with a background in algorithms and optimization; programming skills (e.g. C++) would be a plus.

Supervisor : Solon Pissis (CWI) and Michelle Sweering (CWI)

Keywords : algorithms, graph algorithms, combinatorial optimization, data mining