A STUDY OF CONVEX HULLS INTERSECTION IN CLUSTERING WITH CARDINALITY CONSTRAINTS

AGUSTI SOLANAS

Key words: Convex hulls, Clustering, Statistical Disclosure Control.

Statistical disclosure control (SDC) seeks to transform data in such a way that they can be publicly released whilst preserving data utility and statistical confidentiality. Controlling statistical disclosure is specially important when the protected data belong to individual respondents or entities whose privacy could be put in jeopardy. This kind of data is known as micro-data and one of the most popular techniques for protecting micro-data is micro-aggregation.

Given a data set D, the micro-aggregation problem consists of two steps: (i) generate subsets of D such as the homogeneity of the subsets is maximised and the cardinality of each subset is at least k, (ii) compute the centroid of each subset and replace the original elements in D by the centroid of the subset to which they belong. Multivariate micro-aggregation can be seen as a clustering problem with constraints in the size of the clusters. Each cluster (*i.e.* a set of points in $\mathbb{R}^n$) can be wrapped by a convex hull.

In this article we study the intersection of the convex hulls that wrap the elements of D. We observe that the existence of such intersections lead to a poor within-group homogeneity in terms of sum of square errors (SSE).

CRISES Research Group, UNESCO Chair in Data Privacy, Dept. Computer Science and Mathematics, Rovira i Virgili University, Tarragona, Catalonia, Spain,

E-mail address: agusti.solanas@urv.cat