Hyperbolic Smoothing Method: A Novel Approach for Solving Clustering Problems

Adilson Elias Xavier - UFRJ

Abstract

Clustering analysis can be defined according to numerous criteria, through different mathematical formulations. The methodology considered by the recently published Patent, number 10,217,056 B2 in USPTO (United State Patent and Trademark Office), deals with the general clustering problem, where the measurement of the quality of a solution is given by the sum of an arbitrary monotonic increasing function applied to the distances of each pattern to its nearest centroid, where the measure of distances can be done following different metrics.

The methodology, called hyperbolic smoothing, has a wider scope, and can be applied to clustering according to distances measured in different metrics, such as Euclidian, Minkowski, Manhattan and Chebychev norms.

By smoothing we fundamentally mean the substitution of an intrinsically non-differentiable two-level problem by a completely differentiable single-level alternative. This is achieved through the solution of a sequence of differentiable sub-problems which gradually approaches to the original problem.

An additional improvement considers the partition of the set of observations into two groups: “data in the frontier” and “data in gravitational regions”. The resulting effect is a desirable substantial reduction of the computational effort necessary to solve the clustering problems.

The talk will consider three clustering formulations:

1 - Among many criteria, the most natural, intuitive and frequently adopted criteria is the minimum sum-of-squares clustering (MSSC);

2 - The minimum sum of distances clustering problem according to the Euclidian metric, which is analogous to the multisource Fermat-Weber location problem;

3 - The minimum sum of distances clustering problem according to the Manhattan metric.

In order to show the distinct performance of the proposed methodologies, a set of computational results obtained by solving large test problems is presented.