Pattern Recognition 43 (2010) 731 – 737

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

The hyperbolic smoothing clustering method

Adilson Elias Xavier

Department of Systems Engineering and Computer Science, Graduate School of Engineering (COPPE), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 20 May 2008
Received in revised form 21 October 2008
Accepted 19 June 2009

Keywords:
Cluster analysis
Min-sum-min problems
Nondifferentiable programming
Smoothing

ABSTRACT

The minimum sum-of-squares clustering problem is considered. The mathematical modeling of this problem leads to a *min-sum-min* formulation which, in addition to its intrinsic bi-level nature, has the significant characteristic of being strongly nondifferentiable. To overcome these difficulties, the resolution, method proposed adopts a smoothing strategy using a special C^{∞} differentiable class function. The final solution is obtained by solving a sequence of low dimension differentiable unconstrained optimization subproblems which gradually approach the original problem. The use of this technique, called hyperbolic smoothing, allows the main difficulties presented by the original problem to be overcome. A simplified algorithm containing only the essentials of the method is presented. For the purpose of illustrating both the reliability and the efficiency of the method, a set of computational experiments was performed, making use of traditional test problems described in the literature

1. Introduction

Cluster analysis deals with the problems of classification of a set of patterns or observations, in general represented as points in a multidimensional space, into clusters, following two basic and simultaneous objectives: patterns in the same clusters must be similar to another (homogeneity objective) and different from patterns of other clusters (separation objective), see Hartingan [1] and Späth [2].

Clustering is an important problem that appears in the broadest spectrum of applications, whose intrinsic characteristics engender many approaches to this problem, see Dubes and Jain [3], Jain and Dubes [4] and Hansen and Jaumard [5].

In this paper, a particular clustering problem formulation is considered. Among many criteria used in cluster analysis, the most natural, intuitive and frequently adopted criterion is the minimum sum-of-squares clustering (MSSC). This criterion corresponds to the minimization of the sum-of-squares of distances of observations to their cluster means, or equivalently to the minimization of within-group sum-of-squares. It is a criterion for both the homogeneity and the separation objectives, as, according to the Huygens theorem, minimizing the within-cluster inertia of a partition (homogeneity within the cluster) is equivalent to maximizing the between-cluster inertia (separation between clusters).

The minimum sum-of-squares clustering (MSSC) formulation produces a mathematical problem of global optimization. It is both a nondifferentiable and a nonconvex mathematical problem, with a large number of local minimizers. It is one of the problems in the NP-hard class [6].

In the cluster analysis scope, algorithms use, traditionally, two main strategies: hierarchical clustering methods and partition clustering methods [5,7]. Hierarchical methods, essentially heuristic procedures, produce a hierarchy of partitions of the set of observations according to an agglomerative strategy or to a divisive one. In the former case, the general algorithm starts from an initial partition, in which each cluster contains one pattern, and successively merges two clusters on the basis of a similarity measure until all patterns are in the same cluster. In the latter case, the general algorithm starts from an initial partition with all patterns in the same cluster and, by successive bipartitions, reaches a partition in which each cluster contains one single pattern. In both strategies, the best partition is chosen, by a suitable criterion, from the hierarchy of partitions obtained.

Partition methods, in general, assume a given number of clusters and, essentially, seek the optimization of an objective function measuring the homogeneity within the clusters and/or the separation between the clusters. Heuristic algorithms of the exchange type as the traditional *k*-means algorithm [8] and variations thereof [2,9] are frequently used to find a local minimum of the objective function. However, any mathematical programming technique can be applied to solve the global optimization problem: dynamic programming

E-mail address: adilson@cos.ufrj.br

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.06.018

[10], branch and bound [11], interior point algorithms [12], bilinear programming [13], all kinds of metaheuristics (for instance, see [14,15]) and nonsmooth optimization [16].

The core focus of this paper is the smoothing of the *min-sum-min* problem engendered by the modeling of the clustering problem. In a sense, the process whereby this is achieved is an extension of a smoothing scheme, called hyperbolic smoothing, presented in Santos [17] for nondifferentiable problems in general, in Chaves [18] for the min–max problem and, more recently, in Xavier and Oliveira [19] for the covering of plane domains by circles. This technique was developed through an adaptation of the hyperbolic penalty method originally introduced by Xavier [20].

By smoothing we fundamentally mean the substitution of an intrinsically nondifferentiable two-level problem by a C^∞ differentiable single-level alternative. This is achieved through the solution of a sequence of differentiable subproblems which gradually approaches the original problem. In the present application, each subproblem, by using the implicit function theorem, can be transformed into a low dimension unconstrained one, which, owing to its being indefinitely differentiable, can be comfortably solved by using the most powerful and efficient algorithms, such as conjugate gradient, quasi-Newton or Newton methods.

Although this paper considers the particular MSSC problem, it must be emphasized that the proposed methodology, hyperbolic smoothing, can be used for solving other clustering problem formulations as well.

This work is organized in the following way. A step-by-step definition of the clustering problem, directly connected to the presentation of the proposed hyperbolic smoothing approach, is presented in the next section. The new methodology is described in Section 3. The algorithm and the illustrative computational results are presented in Sections 4 and 5. Brief conclusions are drawn in Section 5.

2. The clustering problem as a min-sum-min problem

Let $S=\{s_1,\dots,s_m\}$ denote a set of m patterns or observations from an Euclidean n-space to be clustered into a given number q of disjoint clusters.

To formulate the original clustering problem as a *min-sum-min* problem, we proceed as follows. Let $x_i, i=1,\dots,q$ be the centroids of the clusters, where each $x_i \in \mathbb{R}^n$. The set of these centroid coordinates will be represented by $X \in \mathbb{R}^{nq}$. Given a point s_j of S, we initially calculate the distance from s_j to the center in X that is nearest. This is given by

$$z_j = \min_{x_i \in X} \|s_j - x_i\|_2. \tag{1}$$

The most frequent measurement of the quality of a clustering associated to a specific position of q centroids is provided by the sum-of-squares of these distances

$$D(X) = \sum_{j=1}^{m} z_j^2. \tag{2}$$

The optimal placing of the centroids must provide the best quality of this measurement. Therefore, if X^* denotes an optimal placement, then the problem is

$$X^* = \operatorname*{argmin}_{X \in \mathbb{R}^{nq}} D(X), \tag{3}$$

where X is the set of all placements of the q centroids. Using (1)–(3), we finally arrive at

$$X^* = \operatorname*{argmin}_{X \in \mathbb{R}^{nq}} \sum_{j=1}^{m} \min_{x_i \in X} \|s_j - x_i\|_2^2. \tag{4}$$

3. Transforming the problem

Problem (4) above can be formulated equivalently as

$$\begin{aligned} &\text{minimize} && \sum_{j=1}^{m} z_j^2 \\ &\text{subject to} && z_j = \min_{i=1,\dots,q} \|s_j - x_i\|_2, \quad j=1,\dots,m. \end{aligned} \tag{5}$$

Considering its definition, each z_j must necessarily satisfy the following set of inequalities:

$$z_j - \|s_j - x_i\|_2 \le 0, \quad i=1,\dots,q. \tag{6}$$

Substituting these inequalities for the equality constraints of problem (5), the relaxed problem becomes

$$\begin{aligned} &\text{minimize} && \sum_{j=1}^{m} z_j^2 \\ &\text{subject to} && z_j - \|s_j - x_i\|_2 \le 0, \quad j=1,\dots,m, \ \ i=1,\dots,q. \end{aligned} \tag{7}$$

Since the variables z_j are not bounded from below, the optimum solution of the relaxed problem will be $z_j=0$, $j=1,\dots,m$. In order to obtain the desired equivalence, we must, therefore, modify problem (7). We do so by first letting $\varphi(y)$ denote $\max\{0,y\}$ and then observing that, from the set of inequalities in (7), it follows that

$$\sum_{i=1}^{q} \varphi(z_j - \|s_j - x_i\|_2) = 0, \quad j=1,\dots,m. \tag{8}$$

For fixed j and assuming $d_1 < \cdots < d_q$ with $d_i = \|s_j - x_i\|_2$, Fig. 1 illustrates the first three summands of (8) as a function of z_j.

Using (8) in place of the set of inequality constraints in (7), we would obtain an equivalent problem maintaining the undesirable property that z_j, $j=1,\dots,m$ still has no lower bound. Considering, however, that the objective function of problem (7) will force each z_j, $j=1,\dots,m$, downward, we can think of bounding the latter variables from below by considering "$>$" in place of "$=$" in (8) and considering the resulting "non-canonical" problem

$$\begin{aligned} &\text{minimize} && \sum_{j=1}^{m} z_j^2 \\ &\text{subject to} && \sum_{i=1}^{q} \varphi(z_j - \|s_j - x_i\|_2) > 0, \quad j=1,\dots,m. \end{aligned} \tag{9}$$

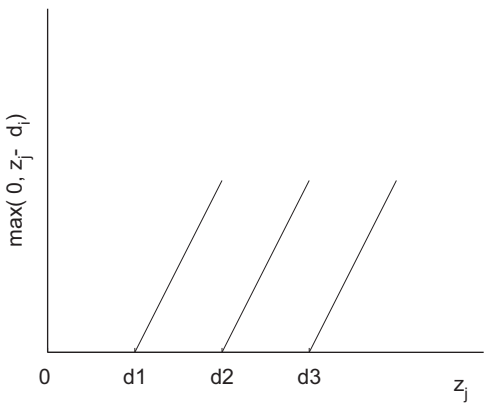

Fig. 1. Summands in (8).