

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

An Iterative Projective Clustering Method

Renata Avros^a, Zakharia Frenkel^{a*}, Dvora Toledano-Kitai^a and Zeev Volkovich^a

^a*Ort Braude College of Engineering, Karmiel 21982, Israel*

Abstract

In this article we offer an algorithm recurrently divides a dataset by search of partitions via one dimensional subspace discovered by means of optimizing of a projected pursuit function. Aiming to assess the model order a resampling technique is employed. For each number of clusters, bounded by a predefined limit, samples from the projected data are drawn and clustered through the EM algorithm. Further, the basis cumulative histogram of the projected data is approximated by means of the GMM histograms constructed using the samples' partitions. The saturation order of this approximation process, at what time the components' amount increases, is recognized as the "true" components' number. Afterward the whole data is clustered and the densest cluster is omitted. The process is repeated while waiting for the true number of clusters equals one. Numerical experiments demonstrate the high ability of the proposed method.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of KES International

Keywords: Projective Clustering; Projected Pursuit; Image Segmentation

1. Introduction

General clustering procedures applied for high-dimensional data classification are frequently based on the Gaussian Mixture Model (GMM). Such model can expose an unsteady performance when the size of the considered dataset is overly small compared to the number of parameters to estimate or when an expected partition is composed

* Corresponding author. Tel.: +972-4-990-1764; fax: +972-4-990-1852.
E-mail address: zfrenkel@braude.ac.il

from clusters with significantly diverse sizes. To avoid these difficulties a projection clustering approach can be applied aiming to find a balance between the parameters' quantity and the generality of the replica. Although the unrelated characteristics of the projected set may actually "hide" the clusters by imaging two items belonging to the same cluster observe as dissimilar as an arbitrary couple of items. Likewise, items could cluster inversely in varying subspaces.

The projected clustering concept presumes that consequential partition can be discovered by projecting onto subspaces of lower dimensionality. Practically the most of existing projected clustering algorithms (see e. g. ¹⁻⁵) are definitely based on an assumption that underlying clusters are depicted by areas of the data of high density separated by sparse areas. This fact is expressed by separated "picks" or "islands" in a subspace corresponding to the overall density of the full space. Hence, the seeking of attractive cluster structure in the high dimensional space can be altered by a corresponding procedure in lower dimensional subspaces. Formally speaking, interestingness is measured by a distance between the distribution of the projected data and a distribution of recognized as uninteresting, which is typically suggested to be normal. So, (see ⁶⁻⁷) any test statistic for testing non-normality (or departure from normality) might be applied as a projection index, quantifying the "interestingness".

In this article we offer a one-dimensional projection pursuit algorithm in the framework of the general Gaussian Mixture Model (GMM) supposing that each cluster is represented by a Gaussian probability density. Note that each projection of GMM distributed data is also GMM distributed. The parameters of GMMs are mostly estimated by the well-known Expectation Maximization Algorithm (EM) finding a maximum likelihood solution. A weakness of this fitting method consists of poor functioning, once high-dimensional data are operated as a large sample size is required in order to attain the required precision.

We propose here a hierarchical projective clustering approach in the spirit of the mentioned earlier projection pursuit perspective based on searching of appropriate one-dimensional subspaces (directions). Note that such an approach is naturally connected to the color space optimizations (see, e.g. ⁸) where an image transformation is constructed in a way that saves as much of the information as possible from the source space though remaining as authentic as possible to the natural mapping. Actually, such a transformation is appearing in our approach as a weighted sum of the three linear-intensity values with the weights evaluated via the clustering projections goodness. The algorithm recurrently divides a dataset by search of partitions via one dimensional subspace discovered by means of optimizing of a projected pursuit function. Aiming to assess the model order (the components' quantity) a resampling technique is employed. For each number of clusters, bounded by a predefined limit, samples from the projected data are drawn and clustered through the EM algorithm. Further, the basis cumulative histogram of the projected data is approximated by means of the GMM histograms constructed using the samples' partitions. The saturation order of this approximation process, at what time the components' amount increases, is recognized as the "true" components' number. Afterward the whole data is clustered, by the EM algorithm with this found number of clusters, and the densest cluster is omitted. The process is repeated while waiting for the true number of clusters equals one.

The paper is organized as follows: In section 2 we present proposed method and discuss its ingredients: the Gaussian Mixture model and the closely connected EM algorithm, Criteria for Projections Selecting and the model selection method. The remaining sections are devoted to the numerical experiments consisting of an application of the presented method to image segmentation and to conclusion.

Nomenclature

GMM	Gaussian Mixture Model
EM	Expectation Maximization
KS	Kolmogorov-Smirnov

Download English Version:

<https://daneshyari.com/en/article/489537>

Download Persian Version:

<https://daneshyari.com/article/489537>

[Daneshyari.com](https://daneshyari.com)