Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 747–761
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

# Ascent EM for fast and global solutions to finite mixtures: An application to curve-clustering of online auctions

Wolfgang Jank

*Department of Decision and Information Technologies, The Robert H. Smith School of Business, University of Maryland, USA*

Received 27 March 2006; accepted 28 March 2006
Available online 27 April 2006

**Abstract**

In this paper we propose two new EM-type algorithms for model-based clustering. The first algorithm, Ascent EM, draws its ideas from the Monte Carlo EM algorithm and uses only random subsets from the entire database. Using only a subset rather than the entire database allows for significant computational improvements since many fewer data points need to be evaluated in every iteration. We also argue that one can choose the subsets intelligently by appealing to EMs highly-appreciated likelihood-ascent property. The second algorithm that we propose builds upon Ascent EM and incorporates ideas from evolutionary computation to find the global optimum. Model-based clustering can feature local, sub-optimal solutions which can make it hard to find the global optimum. Our algorithm borrows ideas from the Genetic Algorithm (GA) by incorporating the concepts of crossover, mutation and selection into EMs updating scheme. We call this new algorithm the GA Ascent EM algorithm. We investigate the performance of these two algorithms in a functional database of online auction price-curves gathered from eBay.com.


*Keywords:* Mixture model; Clustering; Monte Carlo EM; Acceleration; Global optimum; Evolutionary computation; Genetic algorithm; Functional data; Online auction; eBay

## 1. Introduction

The Expectation-Maximization (EM) algorithm is a very popular tool for maximizing objective functions. One reason for this popularity is its wide applicability. Another reason is that, in every iteration, it guarantees an increase in the objective function and converges to (at least) a local maximum. The EM algorithm is particularly appealing for model-based clustering which stems from the fact that it naturally handles situations in which some of the information is unobserved. Model-based clustering assumes that the data originate from a finite mixture of populations, but that the cluster membership of each data point is not directly observable. The EM algorithm appeals to this situation by maximizing the joint likelihood of the observed and unobserved data.

Despite its wide-spread popularity, practical usefulness of EM is often limited by computational inefficiency. In fact, one of the most common criticisms is that it converges only at a linear rate. The convergence can be especially slow if the proportion of unobserved-to-observed information is large. Another drawback is that EM makes a pass through all of the available data in every iteration. Thus, if the size of the data set is large, every iteration can be computationally intensive.

*E-mail address:* wjank@rhsmith.umd.edu.

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.03.013

The main reason for this intensity is that EM requires re-evaluation of the so-called "Q-function" in every E-step. Related research has shown, however, that an exact evaluation may not be necessary, at least not in every iteration. In fact, since EM typically takes very large steps in the earlier iterations, a rough approximation may suffice (Wei and Tanner, 1990). We propose a fast implementation of EM which operates only on a subset of the data. Using only a small subset has the advantage of significantly reducing the computational burden since the method has to evaluate fewer data points. This frees computing resources and speeds-up the computing time. Similar ideas have been proposed earlier (e.g. Ng and McLachlan, 2003). One of the key differences to earlier approaches is that we choose samples randomly. A random selection allows for a statistical treatment of some of the key algorithmic components such as estimation of the progress of the method, choice of the sample size, and monitoring of convergence.

While using only a small sample can lead to enormous computational gains, attention has to be paid to the trade-off between computation and accuracy of the results. In fact, the algorithm will not converge to the true solution if the sample size is held constant. One solution is to increase the sample size successively (but see also Ng and McLachlan, 2003, for alternative approaches). Determining the exact amount by which the sample size should be increased at each iteration is a challenging task and is a topic of ongoing research (Booth and Hobert, 1999; Levine and Casella, 2001; Levine and Fan, 2004; Caffo et al., 2005). In this paper we build upon the work of Caffo et al. (2005) and propose the *Ascent EM* algorithm which chooses the sample size intelligently by measuring the extra information that is available in the data sample. Our example shows that this new implementation can lead to significant computational improvements without sacrificing accuracy of the results.

While the first part of this work deals with computational efficiency, the second part addresses EMs global convergence. The EM algorithm is a greedy method in the sense that it is attracted to the solution closest to its starting value. This can be a problem when several sub-optimal solutions exist. The mixture model, for instance, is well-known to feature many local maxima, especially when the number of mixture-components is large. This means that any solution found by EM may not be the best solution. One ad-hoc approach to alleviate this problem is to initialize EM from a variety of different starting values, but this approach can be burdensome if the parameter-space is of high dimension.

While the EM algorithm is a local optimization method by nature, there is a growing literature on methods that promise convergence to a global solution. Different global optimization paradigms exist. One very popular approach is the concept of *evolutionary computation*, and the *genetic algorithm* (GA) in particular. Evolutionary computation is associated with the groundbreaking work of Holland (1975). Evolutionary algorithms find their inspiration from natural selection and survival of the fittest in the biological world. These algorithms weed out poor solutions, and combine good solutions with other good solutions to create new generations of even better solutions.

We propose a new implementation of the EM algorithm that incorporates the ideas of global optimization. This implementation is based on Ascent EM with features borrowed from the GA to overcome local solutions more systematically. We will refer to this new algorithm as the *GA Ascent EM* algorithm. One of the appeals of this new method is that it is rooted within the framework of the EM algorithm and thus shares many of the well-known properties that the field of statistics has grown so accustomed to.

This paper is organized as follows. In Section 2 we describe the mixture model for model-based clustering and the classic form of the EM algorithm. Section 3 introduces the Ascent EM algorithm for fast computation based on random sub-sampling. Then we describe the GA Ascent EM algorithm that borrows ideas from the GA for global optimization of the mixture likelihood. In Section 4 we investigate the performance of the methods for curve-clustering in a functional database of online auctions gathered off eBay's website. We conclude with additional remarks in Section 5.

## 2. Model-based clustering and the classical EM algorithm

### 2.1. Mixture model for clustering

The traditional mixture model for clustering assumes a set of $n$ $p$-dimensional data vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from a mixture of a finite number of $g$ groups in some unknown proportions $\pi_1, \ldots, \pi_g$. Specifically, we assume that the mixture density of the $j$th data point $\mathbf{x}_j$ $(j = 1, \ldots, n)$ can be written as

$$f\left(\mathbf{x}_j; \boldsymbol{\theta}\right) = \sum_{i=1}^{g} \pi_i f_i\left(\mathbf{x}_j; \boldsymbol{\psi}_i\right), \tag{1}$$