

The Summer Undergraduate Research Fellowship (SURF) Symposium
6 August 2015
Purdue University, West Lafayette, Indiana, USA

Model Selection for Gaussian Mixture Models for Uncertainty Qualification

Yiyi Chen, Xuan Liu
Department of Electrical and Computer Engineering, Purdue University
Guang Lin
Department of Mathematics, Department of Mechanical Engineering Purdue University

ABSTRACT

Clustering is task of assigning the objects into different groups so that the objects are more similar to each other than in other groups. Gaussian Mixture model with Expectation Maximization method is the one of the most general ways to do clustering on large data set. However, this method needs the number of Gaussian mode as input(a cluster) so it could approximate the original data set. Developing a method to automatically determine the number of single distribution model will help to apply this method to more larger context. In the original algorithm, there is a variable represent the weight of each cluster. The weight means how the cluster will affect the data set, more precisely, each data point. So the idea is, we first set the number of the clusters to be a big number, then we are going to apply a penalized likelihood method to update the weights, while we are updating other parameters. The cluster will be deleted if its weight is less than a certain number we set. After all the iteration, the number of clusters will be generated, as well as other parameters of Gaussian model. The results from the simulation(MATLAB) shows that the number of the clusters could be generated from the modified method, and the final result of the clustering perform well to demonstrate the original data set. Although the modified algorithm could be used automatically do the whole clustering process, it need further investigation about its accuracy and improve its speed.

KEYWORDS

Gaussian mixture models, Model selection, EM algorithm, Penalized likelihood.