

The Summer Undergraduate Research Fellowship (SURF) Symposium
4 August 2016
Purdue University, West Lafayette, Indiana, USA

Model Selection Using Gaussian Mixture Models and Parallel Computing

Tian Qiu

Department of Mathematics and Electrical and Computer Engineering, Purdue University

Georgios Karagiannis

Department of Mathematics, Purdue University

Guang Lin

Department of Mathematics and Mechanical Engineering, Purdue University

ABSTRACT

In recent years, model selection methods have seen significant advancement, but improvements have tended to be benchmarked on its efficiency. An effective model selection system requires a robust feature extraction module. A model selection system is developed by using Finite Multivariate Generalized Gaussian Mixture Model, which organizes data points into clusters. Clustering is basically to assign data set into different groups based on their similarity. In this model, expectation maximization method is used to calculate the distance from each point to their dummy center point, where center point will be changing with the process of simulation to get the best fitting results. Parallel computing is utilized to accelerate simulation process. The performance of the developed model is studied through experimental evaluation with ten thousands data points and identification accuracy. The system still can be improved by a new algorithm to separate the cluster. Performance evaluations will be investigated and compared.

KEYWORDS

Machine Learning, Gaussian mixture models, Model selection, EM algorithm, Parallel Computing