Recursive Harmony Search Based Classifier Ensemble Reduction

June, 2018

Prajwal Kailas
K Chandrasekaran

Department of Computer Science and Engineering
National Institute of Technology
Karnataka Surathkal, India

Project objective: Improve accuracy, diversity and reduce memory and storage requirement of ensemble classifier by eliminating redundancy.

Abstract

- In recent times classifier ensembles have become a mainstay in data mining and machine learning. The combination of several classifiers generally results in better performance and accuracy as compared to a single classifier.

- This paper discusses using feature selection and in particular employing recursive harmony search to perform classifier ensemble reduction via feature selection.

- The final ensemble classifier will be a reduced set of the original ensemble classifier, while maintaining diversity and accuracy of the original one.

Introduction

- Ensemble learning: A machine learning approach where multiple classifiers are trained to solve the same problem.
- The main objective of ensemble classifiers is to improve the performance of stand alone classifiers.
- Problem: Redundant classifiers and redundant data.
- Consequence:
  - No improvement in accuracy and diversity.
  - Increased training and testing time.
- Therefore the time and data constraints result in a significant overhead to system memory and run time.

Harmony Search

- Harmony search is a meta heuristic algorithm which is based on finding a solution vector to optimize the cost function of a given optimization problem.
- Essentially harmony search is selecting the set of notes (set of features) that produce the best harmony (global optimum).
- The musician is the decision variable, the musician decides whether a note should be included in the harmony or not. The harmony (feature subset) is composed of notes that the musicians decide on.
- Evaluate the feature subsets to identify the best harmony that is produced by the decisions of the musicians.

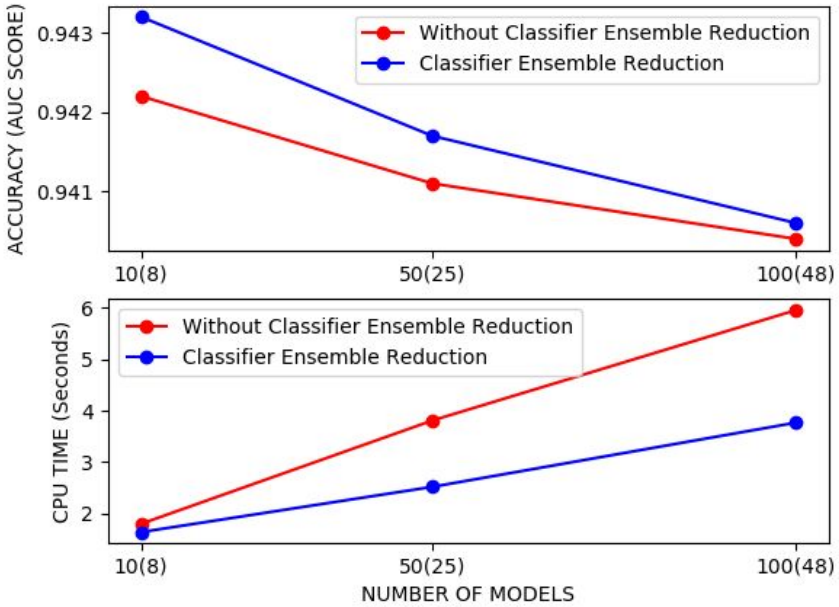
Harmony Search Feature Selection

- The feature selection algorithm to perform classifier ensemble reduction is inspired from the harmony search algorithm.
- The notes are the base classifiers and the harmony is the ensemble classifier.
- The idea is that the musician will select whether to use a base model or not, that is whether the base model should be included in the group of classifiers forming the ensemble classifier.
- Problem: Reliance on a randomly initialized harmony search matrix, which could lead to a suboptimal solution.

Recursive Harmony Search Feature Selection

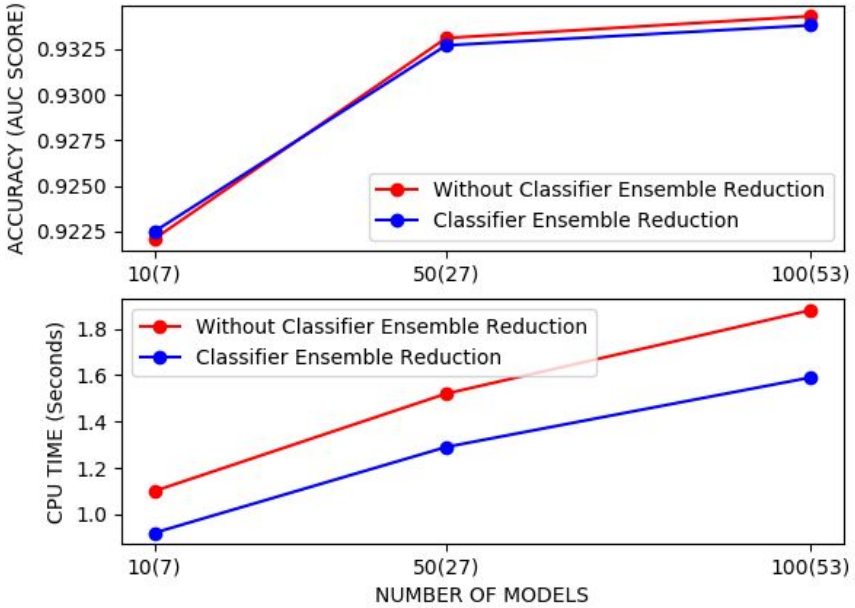
- To overcome the drawback we proposed an extension to the harmony search feature selection algorithm.
- The recursive harmony search feature selection algorithm reduces the reliance of the harmony search feature selection algorithm on the initial matrix.
- Recursive harmony search feature selection initializes multiple harmony matrices and evaluates these matrices recursively (using harmony search) to find the optimal matrix across all the initial matrices.
- This final matrix leads to an optimal solution as the reliance on the randomly initialized harmony search matrix is vastly reduced.

Results: *Accuracy and CPU Run time before and after applying CER.*



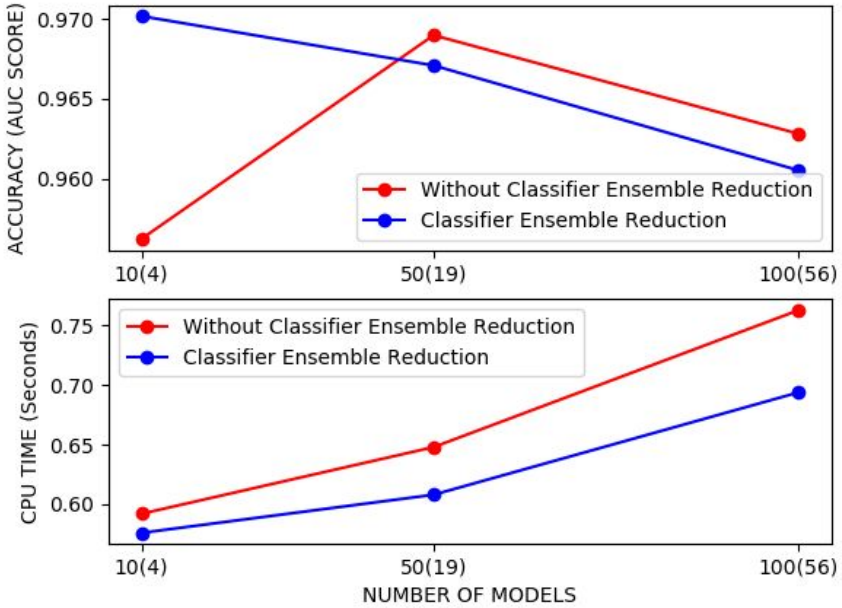
Bank Dataset: 41189 Instances, 21 Features

Results: *Accuracy and CPU Run time before and after applying CER.*



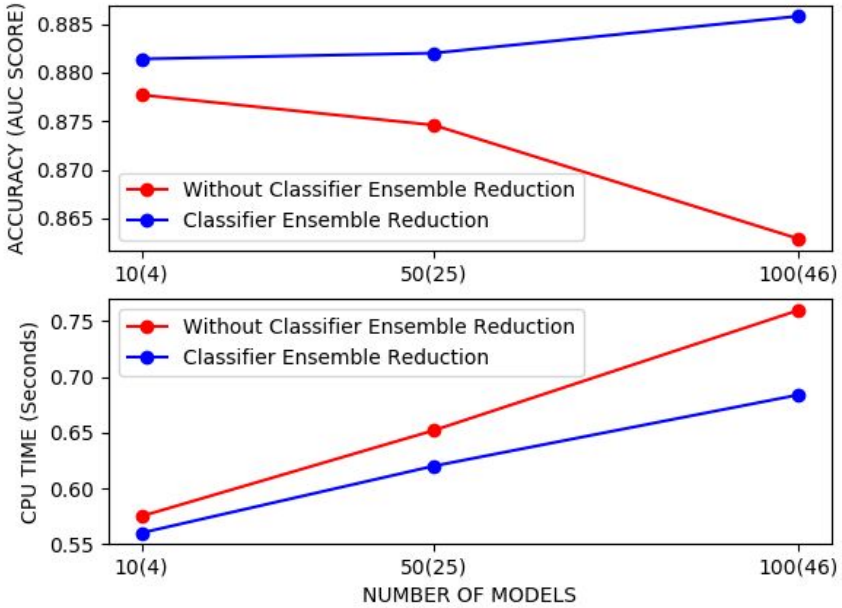
Magic Dataset: 19021 Instances, 11 Features

Results: *Accuracy and CPU Run time before and after applying CER.*



Ionosphere Dataset: 351 Instances, 35 Features

Results: *Accuracy and CPU Run time before and after applying CER.*



Heart Dataset: 270 Instances, 14 Features

Conclusion

- A new robust approach for performing classifier ensemble reduction.
- The recursive classifier ensemble reduction results in a reduced subset of classifiers that are used to train the ensemble classifier. This reduced set of the ensemble classifiers eliminates redundancy and is shown to run faster than the original set of the ensemble classifier.
- The recursive classifier ensemble reduction technique minimizes redundancy, maintains and improves diversity and accuracy, and reduces system memory and run time requirements.