

Feature Selection with Harmony Search for Classification: A Review

Norfadzlan Yusup^{1*}, Azlan Mohd Zain², Nur Fatin Liyana Mohd Rosely², Suhaila Mohamad Yusuf²

¹*Faculty of Computer Science and Information Technology, University Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia*

²*Applied Industrial Analytics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor Darul Takzim, Malaysia*

Keywords: Data Mining, Feature Selection, Nature Inspired Metaheuristic Algorithm, Harmony Search, Classification.

Abstract: In the area of data mining, feature selection is an important task for classification and dimensionality reduction. Feature selection is the process of choosing the most relevant features in a datasets. If the datasets contains irrelevant features, it will not only affect the training of the classification process but also the accuracy of the model. A good classification accuracy can be achieved when the model correctly predicted the class labels. This paper gives a general review of feature selection with Harmony Search (HS) algorithm for classification in various application. From the review, feature selection with HS algorithm shows a good performance as compared to other metaheuristics algorithm such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

1 INTRODUCTION

Data mining is a process of discovering patterns and extracting knowledge from a large set of data. There are various tasks of data mining such as association analysis, anomaly detection, regression, clustering and classification. These data mining tasks can be solved by using a number of different approaches or algorithm (Kotu and Deshpande, 2015). Classification is a data analysis method that extracts models describing important data classes. Such models, called classifiers, predict categorical class labels (Han et al., 2012). Recently, classification using nature inspired metaheuristics algorithms have caught the attention of many researchers.

In the literature, there has been intensifying demand in growth of computational models or methods that motivated by nature inspired or how animals interact and communicate among each other to find food sources. Many optimisation algorithms have been designed and developed by adopting a form of biological-based swarm intelligence. Harmony search (HS) algorithm for example is an optimization algorithm inspired by harmony improvisation process by the musician. There is also a swarm-based algorithm such as Artificial Bee Colony (ABC) that mimics the foraging behaviour of swarm honey bee. Similar to the concept of Ant Colony Optimization

(ACO) and Particle Swarm Optimization (PSO), these type of exploration algorithms is capable of tracing good quality of solutions. Based on Fister et al (2013), all of these algorithms can be named swarm-intelligence based, bio-inspired, physics and chemistry based depending on the sources of inspiration. ACO and PSO are among the most popular swarm-intelligence based algorithms for data mining problems (Martens et al, 2011).

Feature selection methods generally can be categorized into three types which are filter, wrapper and embedded. In addition to these methods, there is a new development of feature selection method such as hybrid method and ensemble method (Ang et al., 2016). The feature selection process is described more details in next section.

For learning and prediction of the models, there are various types of classifiers that have been used with feature selection such as Naïve Bayesian, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision Tree and Artificial Neural Network (ANN).

2 FEATURE SELECTION

Feature selection is a pre-processing techniques that was used to identify relevant features. It is an