

ABSTRACT

Point of sales (POS) data is key data which is used as the evaluation for the store to make improvement. One method to investigate POS data is using data mining technique. Clustering and classification are two major techniques for exploring the hidden structures of data and classifying the unknown data instance, respectively. Due to the nature of difference between unsupervised and supervised methods, clustering and classification are applied separately for different analysis applications depending on whether the data label is available. In this research, a multi-objective framework combining clustering and classification called non-dominated sorting generic algorithm for combining clustering and classification (NSGA-CCC) was proposed to analyze different portions of a dataset. The clustering method is expected to help on rapidly analyzing or identifying the performance measures (Q dataset). The clustering results, labels, are then combined with other information (X dataset) as the inputs of the classification model which classifies the clustering labels by using X dataset. The non-dominated sorting generic algorithm-II (NSGA-II) was integrated in the framework to achieve optimal number of clusters. In this research, hierarchical clustering and decision tree are used for combining clustering and classification as an example. The experimental result shows that the features which influence to the sales are population density, area, department store nearby, and district location. These results can be used as the evaluation from management about the sales performance.

Key words: Clustering, Classification, Multi-Objective Optimization, Non-Dominated Sorting Genetic Algorithm, Point of Sales Data.