FEATURE SELECTION TO IMPROVE THE WRITER IDENTIFICATION FOR OFFLINE WHITEBOARD DATA

Intan Ermahani A. Jalil, Prof. Dr. Siti Mariyam Hj. Shamsuddin
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia
ermahani@utem.edu.my, mariyam@utm.my

ABSTRACT
The process of feature extraction is important to determine the types of generated features in defining the writer identification. Different feature extraction methods might produce different sets of features either from a local or global point of view. However, the extraction of handwriting features has led to high dimensional datasets which resulted in a pool of data. Many of these features are irrelevant or redundant. Unnecessary features increase the size of the search space and make generalization more difficult. The prediction performance of any learning algorithm depends on how efficient the algorithm learns pattern in the data. Due to the high amount of datasets, feature selection is used to improve the writer identification. Feature selection is a process to select features with high significance that can maximize the prediction or classification accuracy. Therefore, this study proposes to use the method of feature ranking and feature combination to find the most significant features for offline whiteboard data. These methods are used to investigate the effect of using them together with feature extraction and discretization method towards improving classification accuracy. This study uses the geometrical feature extraction technique of Higher-Order United Moment Invariant (HUMI) with feature ranking by Grey Relational Analysis (GRA) and linear feature combination together with discretization technique. Based on the experiment, these methods have yielded a promising result that is better than using all features.

KEYWORD
Feature selection, Feature ranking, Feature combination, Writer Identification, Offline whiteboard data