A Hybrid Decision Tree for Printed Tamil Character Recognition Using SVMs

M. Ramanan[1], A. Ramanan[2] and E.Y.A. Charles[2]

[1]Department of Computer Science, Trincomalee Campus, Eastern University, Sri Lanka
[2]Department of Computer Science, University of Jaffna, Jaffna, Sri Lanka

Abstract

Optical character recognition (OCR) is one of the important research areas in image processing and pattern recognition. OCR for printed Tamil text is considered as a challenging problem due to the large number of (i.e., 247) characters with complicated structures and, similarity between characters as well as different font styles. This paper proposes a novel approach for multiclass classification to recognise Tamil characters using binary support vector machines (SVMs) organised in a hybrid decision tree. The proposed decision tree is a binary rooted directed acyclic graph (DAG) which is succeeded by unbalanced decision trees (UDT). DAG implements OVO-based SVMs whereas UDT implements OVA-based SVMs. Each node of the hybrid decision tree exploits optimal feature subset in classifying the Tamil characters. The features used by the decision tree are basic, density, histogram of oriented gradients (HOG) and transition. Experiments have been carried out with a dataset of 12400 samples and the recognition rate observed is 98.80% with the hybrid approach of DAG and UDT SVMs using RBF kernel.

Author Keywords

Tamil OCR, HOG, UDT-SVM, DAG-SVM