Abstract: Novel Nominal Gini-Index Algorithm for Prediction of Major Features from Nominal Dataset: Korea Acute Myocardial Infarction Registry

Heum Park[1], Jeongsu Kim[2], Junehong Kim[3] and Hyuk-Chul Kwon[4*]
[1]Center for U-Port IT Research and Education, [2,3]Division of Cardiology, Department of Internal Medicine & Research Institute for Convergence of Biomedical Science and Technology of Pusan National University Yangsan Hospital, [4]School of Computer Science and Engineering, Pusan National University
{[1]parkheum, [4]hckwon}@pusan.ac.kr, {[2]j25ngsukim, [3]junehongk}@gmail.com

Abstract

This paper presents a novel nominal Gini-Index algorithm for prediction of major features from nominal dataset using the Korea Acute Myocardial Infarction Registry (KAMIR). Only few studies on Gini-Index based nominal feature selection have as yet been completed, and problems in extracting representative features remain for 1) unbalanced dataset for classes, 2) instances having almost all of the features of the datasets, and 3) instances having almost all features with non-null values. To solve these problems, we introduce a novel Gini-Index feature selection algorithm for nominal datasets, and tested the algorithm for prediction of the major features of AMI patients from the KAMIR. In the results, it can show the degrees of importance for features with Gini values and select the major features for given conditions without help by experts.

Acknowledgement

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. 2011-0027460).