

EFFECT OF DISCRETIZATION METHOD ON THE DIAGNOSIS OF PARKINSON'S DISEASE

ERSIN KAYA, OĞUZ FINDIK, İSMAIL BABAOĞLU AND AHMET ARSLAN

Department of Computer Engineering
Faculty of Engineering and Architecture
Selçuk University
Selçuklu, Konya 42075, Turkey
{ ersinkaya; oguzf; ibabaoglu; ahmetarslan }@selcuk.edu.tr

Received April 2010; revised January 2011

ABSTRACT. *Implementing different classification methods, this study analyzes the effect of discretization on the diagnosis of Parkinson's disease. Entropy-based discretization method is used as the discretization method, and support vector machines, C4.5, k-nearest neighbors and Naïve Bayes are used as the classification methods. The diagnosis of Parkinson's disease is implemented without using any preprocessing method. Afterwards, the Parkinson's disease dataset is classified after implementing entropy-based discretization on the dataset. Both results are compared, and it is observed that using discretization method increases the success of classification on the diagnosis of Parkinson's disease by 4.1% to 12.8%.*

Keywords: Parkinson's disease, Entropy-based discretization method, Classification methods

1. **Introduction.** Parkinson's disease is a kind of nervous system disorder which generally arises mostly in men in their 50s. This disease is firstly discovered by James Parkinson, so it was called Parkinson's disease [1]. The symptoms like poverty of movement, slowness of movement, rigidity and rest tremor are commonly diagnosed in patients with Parkinson's disease [2]. Nowadays, no treatment for Parkinson's disease is available. However, if the disease is diagnosed at an earlier time, drug treatments mitigating the effects of the symptoms are implemented at clinic environments [3].

Research into this disease shows that sound distortion occurs on 90% of Parkinson's disease [4,5]. Much research was performed by using voice disorders for the diagnosis of Parkinson's disease [6]. Little et al. used linear discriminant analysis (LDA) to identify the characteristics of sound data to be used in the diagnosis of the disease. For the diagnosis of Parkinson's disease, they composed a model using selected properties with support vector machine (SVM) classifier [7].

The data subjected to preprocessing in the classification process increase the performance of classification [8,9]. Discretization in the data mining is an important preprocessing type. Continuous-valued features in dataset are transformed to discrete values with discretization method. Research shows that discretization of continuous values features increases the performance of the classification. Polat et al. studying to diagnose nerve disease showed that when used with traditional methods like artificial neural network (ANN), least squares support vector machines, and C4.5, discretization increases the performance of classification [8]. Abraham et al. studying on 28 publicly available medical dataset pointed out the effect of discretization on the success of Naïve Bayes classification [10]. Demsar et al. created a predictive model on data consisting of 69 examples and 174 properties belonging to trauma patients. They used decision tree and Naïve Bayes