World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:09, 2016

Improved Classification Procedure for Imbalanced and Overlapped Situations

Authors : Hankyu Lee, Seoung Bum Kim

Abstract : The issue with imbalance and overlapping in the class distribution becomes important in various applications of data mining. The imbalanced dataset is a special case in classification problems in which the number of observations of one class (i.e., major class) heavily exceeds the number of observations of the other class (i.e., minor class). Overlapped dataset is the case where many observations are shared together between the two classes. Imbalanced and overlapped data can be frequently found in many real examples including fraud and abuse patients in healthcare, quality prediction in manufacturing, text classification, oil spill detection, remote sensing, and so on. The class imbalance and overlap problem is the challenging issue because this situation degrades the performance of most of the standard classification algorithms. In this study, we propose a classification procedure that can effectively handle imbalanced and overlapped datasets by splitting data space into three parts: nonoverlapping, light overlapping, and severe overlapping and applying the classification algorithm in each part. These three parts were determined based on the Hausdorff distance and the margin of the modified support vector machine. An experiments study was conducted to examine the properties of the proposed method and compared it with other classification algorithms. The results showed that the proposed method outperformed the competitors under various imbalanced and overlapped situations. Moreover, the applicability of the proposed method was demonstrated through the experiment with real data.

Keywords : classification, imbalanced data with class overlap, split data space, support vector machine
Conference Title : ICDMA 2016 : International Conference on Data Mining and Applications
Conference Location : Tokyo, Japan
Conference Dates : September 05-06, 2016