



Evaluación del algoritmo AR-NSGEP en colecciones de datos desbalanceadas

Evaluation of the AR-NSGEP algorithm in unbalanced datasets

Alain Guerrero Enamorado

Carlos Morell

Sebastián Ventura

Universidad de las Ciencias Informáticas (UCI). Cuba

Universidad Central “Marta Abreu” de las Villas (UCLV). Cuba

Universidad de Córdoba (UCO). España

Resumen

Uno de los grandes problemas que tiene la minería de datos es la existencia del desbalance. Este fenómeno puede afectar gravemente la efectividad de los sistemas de clasificación. Este trabajo persigue como objetivo fundamental obtener información empírica del desempeño del algoritmo AR-NSGEP en colecciones de datos no-balanceados. Se evalúa dicho algoritmo en colecciones de datos con diferentes niveles de desbalance. Se utilizaron colecciones con razones de desbalance entre 1,5 y 40. Durante la etapa de evaluación se utilizaron técnicas de validación cruzada y pruebas estadísticas no-paramétricas para consolidar los resultados obtenidos. La evaluación se realizó con tres métricas muy utilizadas para medir el desempeño en Sistemas Clasificadores con Aprendizaje. Los resultados obtenidos muestran la competitividad del algoritmo AR-NSGEP en colecciones de datos no-balanceados.

Palabras claves: Sistemas Clasificadores con Aprendizaje, Desbalance, Clasificación

Abstract

One of the biggest problems with data mining is the existence of imbalance. This phenomenon can seriously affect the effectiveness of classification systems. The main objective of this work is to obtain empirical infor-



mation of the performance of the AR-NSGEP algorithm in unbalanced datasets. This algorithm is evaluated in datasets with different levels of imbalance. Were used datasets with an unbalance rate between 1,5 and 40. During the evaluation stage, cross-validation techniques and non-parametric statistical tests were used to consolidate the results obtained. The evaluation was carried out with three metrics widely used to measure the performance in Learning Classifier Systems. The obtained results show the competitiveness of the AR-NSGEP algorithm in unbalanced data collections.

Keywords: Learning Classifier Systems, Unbalance, Classification

