A hybrid under-sampling approach for mining unbalanced datasets: applications to banking and insurance

Madireddi Vasu and Vadlamani Ravi*

Institute for Development and Research in Banking Technology,
Castle Hills Road #1, Masab Tank, Hyderabad-500057, (AP) India
Fax: +91-40-23535157
E-mail: madireddi.vasu@gmail.com
E-mail: rav_padma@yahoo.com
*Corresponding author

Abstract: In solving unbalanced classification problems, machine learning algorithms are overwhelmed by the majority class and consequently misclassify the minority class observations. Here, we propose a hybrid under-sampling approach to improve the performance of classifiers. The proposed approach first employs k-reverse nearest neighbour (kRNN) method to detect the outliers from majority class. After removing the outliers, using K-means clustering, K-clusters are selected to further reduce the influence of the majority class. Then, we employed support vector machine (SVM), logistic regression (LR), multi layer perceptron (MLP), radial basis function network (RBF), group method of data handling (GMDH), genetic programming (GP) and decision tree (J48) for classification purpose. The effectiveness of the proposed approach was demonstrated on datasets taken from insurance fraud detection and credit card churn in banking domain. Ten-fold cross validation method was used in the study. It is observed that the proposed approach improved the performance of the classifiers.

Keywords: insurance fraud detection; credit card churn prediction; data mining; unbalanced datasets; machine learning algorithms.

Biographical notes: Madireddi Vasu received his MTech in Information Technology with specialisations in banking technology and information security from the University of Hyderabad, Hyderabad, India and the Institute for Development and Research in Banking Technology (IDRBT), Hyderabad, India. He received his MSc in Mathematics from Pondicherry University, Pondicherry. His research interests include data mining and soft computing.

Vadlamani Ravi is an Associate Professor in IDRBT, Hyderabad. He received his PhD in Soft Computing from Osmania University, Hyderabad & RWTH Aachen Germany. Earlier, he was a faculty at NUS, Singapore for three years. He has published 84 papers in refereed journals/conferences and invited book chapters. He edited *Advances in Banking Technology and Management: Impacts of ICT and CRM* published by IGI Global, USA. He is a Referee for several international journals and on the editorial board of *IJIDS, IJDATS, IJISSS, IJSDS* and *IJITPM*. He is listed in the *Marquis Who's Who in the World* in 2009 and 2010.