Hyperparameters tuning of random forest with harmony search in credit scoring

ABSTRACT

Correct identification of defaulters and non-defaulters in the lending industry is a crucial task for financial institutions. Credit scoring is a tool utilized for credit granting decisions. Recently, Random Forest (RF) is actively researched in credit scoring due to two main benefits, i.e. non-parametric flexibility to account for various data patterns with good classification ability and the computed features importance that can explain the attributes. Hyperparameters tuning is a necessary procedure to ensure good performance of a RF. This paper proposes the use of a metaheuristic, Harmony Search (HS), to form a hybrid HS-RF to conduct hyperparameters tuning. A Modified HS (MHS) is also proposed, forming MHS-RF, for effective yet efficient search of the RF hyperparameters. Along with parallel computing, MHS-RF effectively reduces the computational efforts of the hyperparameters tuning procedure. The proposed hybrid models are benchmarked with standard statistical models on the Lending Club peer-to-peer lending dataset. The computational results show that a well-tuned RF have better performance than statistical models, with MHS-RF reported the best performance yet being the most efficient in hyperparameters tuning of RF.