

Social Networks in Credit Risk Assessment: A Logistic Regression Model

Abstract

This research examines the effects of behavioural aspects on reducing the information asymmetry associated with credit risk assessment. Many lenders have realised the potential of borrowers with thin financial portfolios and have adopted behavioural models as a result. Behavioural data is heterogeneous and collected from various different sources. Therefore, the need for a powerful statistical model that can understand the data and learn from it iteratively is inevitable. Data on the behaviour of borrowers was collected from a European lender via an online repository on 307,000 loan applications given to customers in South East Asia along with their repayment final outcome. The rationale is to classify applicants based on their proximity to like-minded borrowers. A logistic regression model was run on applicants who had ties with bad performers within their social circles to provide evidence of homophily in credit. Logistic Regression model achieved accuracy of 0.57 with a recall of 0.89 on a balanced testing data set. One-tailed independent t-test was run on defaulting and non-defaulting groups where the average number of defaulting friends of the bad borrowers was 0.19 as opposed to 0.14 for the good borrowers. Differences in the group standard errors were significant where 0.0033 for defaulting group while 0.0008 for non-defaulting. The aim is to improve credit scoring using statistical measures derived from social network analysis.

Keywords: Credit risk, financial inclusion, homophily, information asymmetry, logistic regression, social network analysis