

Mining Big Data in statistical systems of the monetary financial institutions (MFIs)

Ashofteh, Afshin

Nova University of Lisbon, Portugal.

Abstract

The financial crisis prompted a number of statutory and supervisory initiatives that require great disclosure by financial firms of their data to a central system. Recently, core banking and payment systems data as a main big data sources of monetary financial Institutions (MFI's) have been used to monitor different kind of risks, however distress situations for MFI's are relatively infrequent events and as the same time under the pressure of rapid changes in compliance and rules. The very limited information for distinguishing dynamic fraud from genuine customer or monetary and financial institution behavior in an extremely sparse and imbalanced big data environment with probable change points in data distribution is making the instant and effective fraud detection and banking Big Data management more and more difficult and challenging. Being still a recent discipline, few research has been conducted on imbalanced classification for Big Data. The reasons behind this are mainly the difficulties in adapting standard techniques to the MapReduce programming style and inner problems of imbalanced data, namely lack of data, small disjuncts and data distribution changes. These are accentuated during the data partitioning to fit the MapReduce programming style and data mining process. This paper is going to summarize some technical problems of imbalanced data and artificial data for the adjustment of big data for MFI's and to investigate how it can be made ready for implementation.

Keywords: *Big Data, Artificial data, Imbalanced classification, Monetary financial institutions.*
