SURVEY ARTICLE

Surveillance on Bigdata to Mine Pattern

N. Monica[1], Dr. K. Ramesh Kumar[2], T. Nelson Gnanaraj[3]

[1]MTech Scholar, Department of Information and Technology
Hindustan University, Chennai, TamilNadu, India
monitech26@gmail.com
[2]Associate Professor, Department of Information and Technology
Hindustan University, Chennai, TamilNadu, India
rameshkumar.dr@gmail.com
[3]Mtech Scholar, Department of Information and Technology
Hindustan University, Chennai, TamilNadu, India
nelsonraj.27@gmail.com

Abstract

Big data is a collection of large amount of data with various types of data and usable to be processed at much higher frequency. One of the most popular knowledge discovery approaches is to find frequent items from a transaction data set and derive association rules. Pattern finding is one of the most computationally expensive steps in large data sets. Patterns often referred to association rules. Association rule plays an important role in the process of mining data for sequential pattern. Association rules are used to acquire interesting rules from large collections of data which expresses an association between items or sets of items. Apriori is a classic algorithm for learning association rules. It is designed to operate on databases containing transactions. Apriori algorithm attempts to find subsets which are common to at least a minimum number C of the item sets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time and groups of candidates are tested against the data. The algorithm terminates when no further Successful extensions are found. In this paper we enhance Apriori algorithm to solve its complexity over large data sets. We first collect variety of data and then integrate both structured and unstructured data using MapReduce to find out sequential pattern from the required data sets.

Keywords: MapReduce; Apriori; Association Rule; Pattern mining; Variety of data

Full Text: http://www.ijcsmc.com/docs/papers/January2014/V3I1201442.pdf