Chapter 2
A Detailed Study on Classification Algorithms in Big Data

Saranya N.
Sri Eshwar College of Engineering, Coimbatore, India

Saravana Selvam
PSR Engineering College, India

ABSTRACT

After an era of managing data collection difficulties, these days the issue has turned into the problem of how to process these vast amounts of information. Scientists, as well as researchers, think that today, probably the most essential topic in computing science is Big Data. Big Data is used to clarify the huge volume of data that could exist in any structure. This makes it difficult for standard controlling approaches for mining the best possible data through such large data sets. Classification in Big Data is a procedure of summing up data sets dependent on various examples. There are distinctive classification frameworks which help us to classify data collections. A few methods that discussed in the chapter are Multi-Layer Perception Linear Regression, C4.5, CART, J48, SVM, ID3, Random Forest, and KNN. The target of this chapter is to provide a comprehensive evaluation of classification methods that are in effect commonly utilized.

DOI: 10.4018/978-1-5225-9750-6.ch002

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

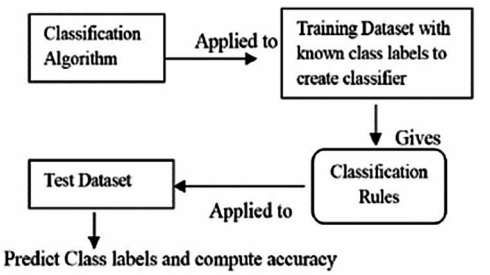
INTRODUCTION

Big data is which grows each year in a system. Invention in research, as well as scientific factors, has impacted the size of data that grows every day with the goal to enhance profitable activities. On the other hand, information accessibility and exploration of thoughts have resulted in the delivery of a completely progressive modification of the internet, which has been utilized in the most beneficial ways. As a result, billions of bytes of information are created every day hence throwing much information at regular intervals. Since newly provided data night is unstructured, structured or possibly confused; we need classification techniques in case of unstructured or complex data. A classification is an approach that allots data in a class to focus on Categories or groups (Owais & Hussein, 2016). The objective of grouping would be to precisely estimate the prospective class of every single file in the data.

Diverse Classification strategies make use of distinctive methods intended for finding the relationships among the list of data units. These types of relationships are generally compacted in a model, which may then have the ability to connect to an alternate dataset when class assignments are covered up. Classification models are broken down through differentiating the expected valuations as well as known values in a wide scope of test data (Revathy & Lawrance, 2017). The target data to get a classification model is normally isolated into several sets of data: one for understanding the genuine model; another for testing the actual model.

The overall classification technique is shown in Figure 1 below. The purpose of the learning stage is to assemble the classification model and testing stage. In the learning stage, a classification strategy is produced to portray a redid set of data classes. In both learning and training phase, the classification approach develops the classifier using a method of comprehension from the training dataset and the related class names.

Figure 1. Classification process



15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-detailed-study-on-classification-algorithms-in-big-data/238603

Related Content

Measuring the Quality of Healthcare Services in Bangladesh
Fahima Khanamand Nayem Rahman (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 15-31).*
www.irma-international.org/article/measuring-the-quality-of-healthcare-services-in-bangladesh/232323

A Novel Approach to Segmentation Using Customer Locations Data and Intelligent Techniques
Baar Öztayi, Ugur Gokdere, Esra Nur Simsekand Ceren Salkin Oner (2017). *Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics (pp. 21-39).*
www.irma-international.org/chapter/a-novel-approach-to-segmentation-using-customer-locations-data-and-intelligent-techniques/170336

Spatial and Temporal Distribution of Pedestrian Crashes
(2019). *Big Data Analytics in Traffic and Transportation Engineering: Emerging Research and Opportunities (pp. 122-151).*
www.irma-international.org/chapter/spatial-and-temporal-distribution-of-pedestrian-crashes/220533

A Comprehensive Review on Leveraging Business Intelligence for Enhanced Marketing Analytics
P. S. Venkateswaran, Divya Marupaka, Sachin Parate, Amit Bhanushali, Latha Thammareddiand P. Paramasivan (2024). *Data-Driven Decision Making for Long-Term Business Success (pp. 34-48).*
www.irma-international.org/chapter/a-comprehensive-review-on-leveraging-business-intelligence-for-enhanced-marketing-analytics/335562

Conceptual View on Healthcare Digitalization: An Extended Thematic Analysis
Robert Furdaand Michal Gregus (2017). *International Journal of Big Data and Analytics in Healthcare (pp. 35-54).*
www.irma-international.org/article/conceptual-view-on-healthcare-digitalization/197440