

Chapter 3

Data Classification: Its Techniques and Big Data

A. Sheik Abdullah

Thiagarajar College of Engineering, India

R. Suganya

Thiagarajar College of Engineering, India

S. Selvakumar

*G. K. M. College of Engineering and Technology,
India*

S. Rajaram

Thiagarajar College of Engineering, India

ABSTRACT

Classification is considered to be the one of the data analysis technique which can be used over many applications. Classification model predicts categorical continuous class labels. Clustering mainly deals with grouping of variables based upon similar characteristics. Classification models are experienced by comparing the predicted values to that of the known target values in a set of test data. Data classification has many applications in business modeling, marketing analysis, credit risk analysis; biomedical engineering and drug retort modeling. The extension of data analysis and classification makes the insight into big data with an exploration to processing and managing large data sets. This chapter deals with various techniques, methodologies that correspond to the classification problem in data analysis process and its methodological impacts to big data.

INTRODUCTION

Data is an abstract concept from which information and knowledge are derived. Raw unprocessed data often moves and crosses stage by stage for its exact representation and processed form of representation. Data is a collection of facts which is the representation of values and measurements.

Meanwhile information is referred to as processed data. It reveals the content or message through direct or indirect form of representation. Hence it is in a meaningful form of representation which can be easily conveyed and understood by the users. It resolves uncertainty and ambiguity.

DOI: 10.4018/978-1-5225-2031-3.ch003

Data Classification

Qualities of Data

The quality signifies the characteristics of data which are specifically suited for the data analysis process. The following characteristics represent the quality of a good data:

1. Accurate
2. Represented numerically
3. Relationship
4. Signified for definite purpose
5. Completeness
6. Clearly Understandable

Types of Data Elements

At the start of every data analysis it is necessary to identify the type of data which can then be considered for analysis. The following represents the types of data elements which can be used up for the determination of the type of data.

Continuous Data

These are the type of data elements which are defined upon an interval scale. Examples include income of employees in an organization, sales of an enterprise and so on.

Categorical Data

These kinds of data elements are of three types:

1. **Ordinal Data:** The type of data elements which takes restricted set of values with meaningful ordering. Example includes the classification of age into young, middle age and old group.
2. **Nominal Data:** These are the type of data elements which takes restricted set of values with no any such meaningful ordering between them. Example includes profession of employees, marital status and so on.
3. **Binary Data:** These are the types of data elements that can take only two values. Examples include gender and employment status of an employee.

Data Standardization

Data standardization is the mechanism of normalizing the data to a defined specified range. It provides the mechanism of coding the data to a smaller specified range. The following are the data normalization procedures used up for scaling the given variable.

16 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/data-classification/178096

Related Content

Ethical Issues and Concerns in Collection of Marketing Information and Marketing Intelligence: Ethical Issue in Collection of Information

Pratap Chandra Mandal (2019). *International Journal of Business Intelligence Research* (pp. 16-28).

www.irma-international.org/article/ethical-issues-and-concerns-in-collection-of-marketing-information-and-marketing-intelligence/232237

Nonlinear Efficiency in DEA Relative to "Ideal Reference"

P. Sunil Dharmapala (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 1637-1647).

www.irma-international.org/chapter/nonlinear-efficiency-in-dea-relative-to-ideal-reference/107354

Analytics for Smarter Buildings

Young M. Lee, Lianjun An, Fei Liu, Raya Horesh, Young Tae Chae and Rui Zhang (2014). *International Journal of Business Analytics* (pp. 1-15).

www.irma-international.org/article/analytics-for-smarter-buildings/107066

Applications of System Dynamics and Big Data to Oil and Gas Production Dynamics in the Permian Basin

James R. Burns and Pinyarat Sirisomboonsuk (2022). *International Journal of Business Analytics* (pp. 1-22).

www.irma-international.org/article/applications-of-system-dynamics-and-big-data-to-oil-and-gas-production-dynamics-in-the-permian-basin/314223

Machine Learning and Deep Learning for Big Data Analysis

Tarun Pradhan, Prathamesh Nimkar and Kavita Jhaharia (2024). *Big Data Analytics Techniques for Market Intelligence* (pp. 209-240).

www.irma-international.org/chapter/machine-learning-and-deep-learning-for-big-data-analysis/336351