INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

DATA MINING TECHNIQUES: A REVIEW

SUNIL SAINI

Assistant Professor, Deptt. of Computer Sc. & Engg., Kurukshetra Institute of Technology & Management, Kurukshetra, Haryana.

Accepted Date: 15/06/2017; Published Date: 01/07/2017

Abstract: - Data mining is very useful in this colossal data world where everyone is generating. It is used for mining useful knowledge, patterns, rules etc for finding useful insights from this high scale data. These mined insights will help business, organization or an individual to earn maximum profit and to optimize their process. This paper provides a brief introduction to various data mining techniques which are useful in finding information from large repositories of information and different applications of data mining. These techniques predict future trends and behavior for making knowledge driven decisions. These techniques find answers of different question that traditionally were time consuming.

Keywords: Association Rule mining, Classification, Clustering, Data Mining, Outlier Analysis

Corresponding Author: SUNIL SAINI

Access Online On:

www.ijpret.com

How to Cite This Article:

Sunil Saini, IJPRET, 2017; Volume 5 (11): 23-26

PAPER-QR CODE

I INTRODUCTION

Data mining is the current topic for every researcher because there is hidden knowledge present in the data. These techniques do not focus on data management, statistics but it mainly focuses on computational aspects of data processing. Data mining is rapidly growing in different disciplines like machine learning, database research, high performance commerce and many more. Data mining discovers useful information and patterns from the data.

II TYPES OF DATA MINING TECHNIQUES

Data mining is very useful in various areas like healthcare, market based analysis, engineering, fraud detection, intrusion detection, banking, bioinformatics and many more [1].

A. CLASSIFICATION

Classification technique is used to find the correct category for the available objects. It is used when categories and collection of data is given. Classification is a kind of mapping function which maps the set of objects to their respective classes. In this technique data is divided into training set and test set. Training set data is used to train the classification algorithm and test set data is used to find the accuracy of classification model.

Various classification models are support vector machine, Bayesian classification, decision tree etc. [2]

B. CLUSTERING

Clustering is a process of dividing similar objects into a number of clusters [3]. In machine learning it is classified as unsupervised learning. Objects in intra clusters are similar to each other in some sense but objects in inter clusters are not similar [4]. Clusters are used to organize, navigate and summarize the data [5]. Clustering is used in many applications like data analysis, document clustering, pattern recognition and many more. Various clustering algorithms are partition clustering, hierarchical clustering, density based clustering and graph based clustering [6].

C. OUTLIER ANALYSIS

An outlier often called an anomaly is a particular data point or, in some instances, a small set of data points that are inconsistent with the rest of the data population [7]. Outlier detection has been used for many decades to detect and remove abnormal data points from data. Outlier detection separates outlier data or abnormal data from normal data using either: abnormality

detection which compares new data to a model of normality (or a model of abnormality); or outlier classification which classifies new data as either normal or abnormal [8]. The original methods were arbitrary but today, principled and systematic techniques are used. These include distance-based; density-based; statistical (including regression); machine learning (including decision trees, expert systems and clustering); information theory; spectral decomposition; neural networks; support vector machines (SVMs); and, natural computation derived from artificial immune systems [9].

D. ASSOCIATION RULE MINING

Association mining is used to find relationship between different items from large data sets. An association rule is expressed as X-> Y, where X and Y are sets of items. This rule mining helps the organization to take certain decisions like to decrease particular item or to increase particular item by analyzing customer shopping behavior, catalogue design etc. [2]. Each association is in the form of [X → Y, sup, conf] where support and confidence is used to notify frequent items in transactions and item likely to be occur in transactions respectively [10]

Various algorithms for association rule mining are AIS algorithm, SETM algorithm, Apriori algorithm, FP-growth algorithm etc. [11]

III CONCLUSION

Data mining is an area of colossal data and information, but to extract information many different techniques are used like classification, clustering, rule mining and outlier analysis. With this paper researchers are able to know the basics of these techniques and methods under these techniques. With these techniques useful patterns, information, insight can be gained which can be helpful for an individual, an organization or to a business.

REFERENCES

1. http://bigdata-madesimple.com/14-useful-applications-of-data-mining/
2. Bharati M. Ramageri et al., “Data Mining Techniques and Applications “, Indian Journal of Computer Science and Engineering, ISSN : 0976-5166
3. Zhao Jie etal.,” Clustering Using Improved Cuckoo Search Algorithm”, Springer 2014
4. Maulik Ujjwal et al.,” Genetic algorithm-based clustering technique”, Pattern Recognition Society. Elsevier, 2000
5. P. Manikandan et al.,” Data Clustering using Cuckoo Search Algorithm”, Springer, 2014
6. Saida Ishak et al.,”*A* New Algorithm for Data Clustering Based on CuckooSearch Optimization”, Springer, 2014

7. Chandola Varun et al. ," Anomaly Detection: A Survey", ACM Computing Surveys, Vol. 41, No. 3, Article 15, July 2009

8. Hossein Moradi et al.," Outlier Detection in Stream Data by Machine Learning and Feature Selection Methods", International Journal of Advanced Computer Science and Information Technology, 2013

9. Victoria J. Hodge et al., "A survey of Outlier Detection Methodologies", White Rose Consortium e Prints Repository, 2004

10. Feng Tao et al. ,"Weighted Association Rule Mining using Weighted Support and Significance Framework", Proc. Ninth ACM SIGKDD, 2003

11. Trupti A. Kumbhare et al.," An Overview of Association Rule Mining Algorithms", International Journal of Computer Science and Information Technologies, 2014