# ALGORTHMIC APPROACHES IN DATA MINING

[1]R. Sathishkumar, [2]Selvakumar, [3]S. Subathra

[1][2][3]Assistant Professor in Computer Science

Dhanalakshmi Srinivasan College of Arts and Science for Women (Autonomous)

Perambalur

## ABSTRACT

Data mining may be a process which finds useful patterns from great deal of knowledge . The paper discusses few of the info mining techniques, algorithms and a few of the organizations which have adapted data processing technology to enhance their businesses and found excellent results. Research on data processing has successfully provided the use of various tools, methods, methods and approaches for various purposes and problem solving.data processing has become an integral a part of many application domains like data ware housing, predictive analytics, business intelligence, bio-informatics and decision support systems. Prime objective of knowledge mining is to effectively handle large scale data, extract actionable patterns, and gain insightful knowledge. data processing is a component and parcel of datadiscovery in databases (KDD) process. Success and improved deciding normally depends on how quickly one can discover insights from data. These insights may not be able to execute optimal actions, they may be used in operational processes and may even predict future behavior.This paper presents an summary of varied algorithms necessary for handling large data sets. These algorithms define the various structures and methods implemented to handle large data.The review also discusses the overall strengths and limitations of those algorithms. This paper can quickly guide or be an eye fixed opener to the info mining researchers on which algorithm(s) to pick and apply in solving the issues they're going to be investigating.

**KEYWORDS:**

Data processing Techniques; data processing algorithms; data processing applications.

## 1. INTRODUCTION

A number of definitions about data processing are laid forth by various researchers. Some have defined data processing as a process of discovering useful or actionable knowledge in large scale data. consistent with Zaki and Meiradataminingis that the process of discovering insightful, interesting, and novel patterns, also as descriptive, understandable, and predictive models from large-scale data. Another definition of knowledge mining as coined by Ozer and Garcia et.al.is that the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of knowledge .

Data processing also means knowledge discovery from data which describes the standard process of extracting useful information from data . As acknowledged by Kamruzzaman, Haider and Hasan ,many of us treat data processing as a synonym for an additional popularly used term, Knowledge Discovery in Databases, or KDD. This observation is quiet true if one can closely check out the interpretations which are made by several researchers about datamining. However as acknowledged by Kamruzzaman, Haider and Hasan ,data processing is additionally treated simply as an important step within the process of data discovery in databases.

Data mining may be a logical process that's wont to search through great deal of knowledge so as to seek out useful data. The goal of this system is to seek out patterns that were previously unknown. Once these patterns are found they will further be wont to make sure decisions for development of their businesses. Three steps involved are,

- Exploration
- Pattern identification
- Deployment

Exploration: within the initiative of knowledge exploration data is cleaned and transformed into another form, and important variables then nature of knowledge supported the matter are determined. Pattern Identification: Once data is explored, refined and defined for the precise variables the second step is to make pattern identification. Identify and choose the patterns which make the simplest prediction. Deployment: Patterns are deployed for desired outcome.

## 2. DATA PROCESSING ALGORITHMS AND TECHNIQUES

Various algorithms and techniques such as classification, clustering, regression, synthesisIntelligence, neural networks, association rules, end trees, genetic algorithm, nearby neighborsMethod etc. are used for knowledge extraction from databases.A data mining algorithm may be a set of heuristics and calculations that makes a knowledge mining model from data.It are often a challenge to settle on the acceptable or best suited algorithm to use to unravel a particular problem. albeitone can use different algorithms to perform an equivalent tasks, each algorithm yield a special set of results, and a few algorithms can even produce quite one sort of results. Some algorithms can perform classification process, that is, they will predict one or more discrete variables, supported the opposite attributes within the data set. Some algorithms perform regression purposes, they will predict more or continuous variables supported the opposite attributes within the data set.

Asacknowledged by Microsoft , some algorithms can perform segmentation,they divide data into groups,or clusters of things that have similar properties.While some algorithms are often associative by finding correlations between different attributes during a set, some are often used for sequence analysis processes, that's they will be wont to summarise sequence or episodes in data, such as an internet path flow [26]. However, all of the aforementioned sorts of algorithms are often categorized into two large categories: Supervised learning and Unsupervised learning algorithms.

The following sub-sections briefly discuss the 2 categories: supervised and unsupervised learning. Several samples of some of the aforementioned algorithms in each of the said categories also are given as a summary in Table 1 and a couple of . Basically both Table 1 and a couple of show a general discussion of a number of the strengths and limitations of a number of these algorithms.

## 2.1. Classification

Classification is that the most ordinarily applied data processing technique, which employs a group of pre-classified examples to develop a model which will classify the population of records at large. Fraud detection and creditriskapplications are particularly compatible to the present sort of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. the info classification process involves learning and classification. Learning data is analyzed by the classification method. Inclassification test data are wont to estimate the accuracy of the classification rules. If the accuracy is acceptable the principles are often applied to the new data tuples. For a fraud detection application, this is able to Includecomplete records of fraud and valid acts determined on a record-by-record basis.

The classifier-training algorithm uses these pre-classified examples to work out the set of parameters required for correct discrimination. The algorithm encodes these parameters in a model called aclassifier.Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification supported Associations

## 2.2. CLUSTERING

Clustering are often said as identification of comparable classes of objects. By using clustering techniques we will further identify dense and sparse regions in object space and may discover overall distribution pattern and correlations among data attributes. Classification approach also can be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering are often used as preprocessingapproach for attribute subset selection and classification. for instance , to make group of consumers supported purchasing patterns, to categories genes with similar functionality.

Types of clustering methods

- Partitioning Methods
- HierarchicalAgglomerative (divisive) methods
- Density based methods
- Grid-based methods

- Model-based methods

### 2.3. Predication

Regression technique are often adapted for predication. Multivariateanalysis are often wont to model the relationship between one or more independent variables and dependent variables. In data processing independent variables are attributes already known and response variables are what we would like to predict. Unfortunately, many real-world problems aren't simply prediction. as an example , sales volumes, stock prices, and merchandise failure rates are all very difficult to predict because they'lldepend upon complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) could also be necessary to forecast future values.an equivalent model types can often be used for both regression and classification. for instance , the CART (Classification and Regression Trees) decision tree algorithm are often wont to build both classification trees (to classify categorical response Variables) and regression trees (predict continuous response variables). Neural networks can also create both classification and regression models.

Types of regression methods

- Rectilinear regression
- Multivariate rectilinear regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

### 2.4. Association rule

Association and correlation is typically to seek out frequent item set findings among large data sets. this sort of finding helps businesses to form certain decisions, like catalogue design, cross marketing and customershopping behavior analysis. Association Rule algorithms got to be ready to generate rules confidently values but one. However the amount of possible Association Rules for a given dataset is usually very large and a high proportion of the principles are usually of little (if any) value.

Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

### 2.5. Neural networks

Neural network may be a set of connected input/output units and every connection features a weight present with it. During the training phase, network learns by adjusting weights so onbe ready to predict the right classlabels of the input tuples. Neural networks have a significant ability to extract material from complex onesor imprecise data and may be wont to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are compatible for continuous valued inputs andoutputs. For instance handwritten character reorganization, for training a computer to pronounce Englishtext and lots of world business problems and have already been successfully applied in many industries. Neural networks are best at identifying patterns or trends in data and compatible for prediction or forecasting needs.Types of neural networks Back Propagation.

## 3. Data Processing Applications

Data mining may be a relatively new technology that has not fully matured. Despite this, there are variety of industries that are already using it on a daily basis. a number of these organizations include retail stores, hospitals, banks, and insurance companies. Many of those organizations are combining data processing withSuch as statistics, system authentication and other important tools.data processing are often wont to find patterns and connections that might rather be difficult to seek out. This technology is fashionable manybusinesses because it allows them to find out more about their customers and make smart marketing decisions. Here is overview of business problems and solutions found using data processing technology.

### 3.1.FBTODutch insurance firm Challenges

• to scale back spam costs.

• Increase efficiency of selling campaigns.

• Increase cross-selling to existing customers,

Using inbound channels like the company's sell centerand the internet a 1 year test of the solution's effectiveness. Results

• Provided the marketing team with the power to predict the effectiveness of its campaigns.

• Increased the efficiency of selling campaign creation, optimization, and execution. • Decreased mailing costs by 35 percent. • Increased conversion rates by 40 percent.

### 3.2. ECtel Ltd., Israel Challenges

• Fraudulent activity in telecommunication services. Results

• Significantly reduced telecommunications fraud for quite 150 telecommunication companies worldwide.

• Saved money by enabling real-time fraud detection.

### 3.3. Provident Financial's Home credit Division, ukChallenges

• No system to detect and stop fraud. Results Reduced frequency and amount of agent and customer fraud.

• Saved money through early fraud detection. • Saved investigator's time and increased prosecution rate.

### 3.4. Standard Life Mutual Financial Services Companies Challenges

• Identify the key attributes of clients interested in their mortgage offer. Cross sells Standard Standard Life Bank products to customers of other Standard Life companies.

• Develop a remortgage model which might be deployed on the group internet site to look at the Profit from the mortgage business is accepted by Standard Life Bank.

## 4. Conclusion

Data mining has importance regarding finding the patterns, forecasting, discovery of

data etc., in several business domains. data processing techniques and algorithms like classification, clusteringetc., helps find the patterns to make a decision upon the longer term trends in businesses to grow. data processing haswide application domain almost in every industry where the info is generated that’s why data processing is considered one among the foremost important frontiers in database and knowledge systems and one among the foremost promising interdisciplinary developments in Information Technology. thanks to the rise within the amount of knowledge coming from everywhere(online: blogging,social media,databases,etc.), it hasbecome difficult to handle the info , to seek out associations, patterns and to analyse the massive data sets. Consequently, large Many technologies are being developed to extract meaningful data from large collections of text datausing different text mining techniques. Different tools, algorithms and methods which are getting used to mine and analysethe data, perform differently on the info collections as has been indicated within the review which has been made during this
paper. Choosing the simplest algorithm to use for a selected analytical task are often a challenge. While you'll use different algorithms to perform an equivalent business task, each algorithm produces a special result, and a few algorithms can produce quite one sort of result.

**References**

1. Jiawei Han and MichelineKamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman,2nd ed.

2. Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM.

3. Crisp-DM 1.0 Step by step Data Miningguidefrom http://www.crisp-dm.org/CRISPWP-0800.pdf.

4. Customer Successes in your industry from http://www.spss.com/success/?source=homepage&hpzone=nav_bar.

5.https://www.allbusiness.com/Technology /computer-software-data-management/ 633425-1.html, last retrieved on15th Aug 2010.

6. http://www.kdnuggets.com/.