Classification of data mining techniques with its applications

Aditya Vishwakarma

Aadi2186@yahoo.co.in

Shri ram group of institution

Department of Computer Science & Engineering

Abstract

Data mining is a method that analyzes massive amounts of data to uncover insightful patterns. The paper examines a few data mining methods, algorithms, and enterprises that have used data mining technologies to enhance their operations with great success.

Keywords: Data mining Techniques; Data mining algorithms; Data mining applications.

1. Overview of data mining

Large databases and vast amounts of data have been produced in many different fields as a result of the growth of information technology. A technique to storing and manipulating this priceless data for future decision-making has emerged as a result of study in databases and information technology. Data mining is the process of removing relevant information and patterns from massive amounts of data. Data/pattern analysis, knowledge mining from data, knowledge extraction, and knowledge discovery process are some names for it.



Figure 1. Knowledge Discovery model

In order to find meaningful data from a big volume of data, a method known as data mining is utilized. This method seeks to identify previously undiscovered patterns. Once these patterns have been identified, they can then be used to guide specific business development decisions.

Three steps involved are

* Exploration
* Pattern identification
* Deployment

Exploration: Data cleaning and transformation are the first steps in data exploration. Next, significant factors and the type of the data based on the issue are identified.

Pattern identification: Pattern identification is the second phase after data has been examined, clarified, and specified for the particular variables. Find and select the patterns that offer the most reliable predictions.

Deployment: Patterns are used to get the desired result.

## Data Mining Algorithms and Techniques

Knowledge discovery from databases uses a variety of algorithms and techniques, including classification, clustering, regression, artificial intelligence, neural networks, association rules, decision trees, genetic algorithms, and the nearest neighbor method, among other.

## Classification

The most widely used data mining technique is classification, which uses a set of previously categorized instances to create a model that can classify the majority of records. Applications for credit risk and fraud detection are especially well suited to this kind of investigation. This method usually uses classification algorithms based on neural networks or decision trees. Learning and classification are both involved in the data classification process. The training data are examined by a classification algorithm in learning. Data from classification tests are used to gauge how accurate the rules are. The rules can be applied to the new data tuples if the accuracy is deemed acceptable. This would comprise comprehensive records of both valid and fraudulent actions identified record-by-record for a fraud detection application. the training of the classifier.

*Types of classification models:*

* Classification by decision tree induction
* Bayesian Classification
* Neural Networks
* Support Vector Machines (SVM)
* Classification Based on Associations

## Clustering

Identification of comparable classes of objects is referred to as clustering. We can further define dense and sparse regions in object space and learn about the general distribution pattern and relationships between data attributes by employing clustering techniques. The classification strategy can also be used to distinguish between groups or classes of objects, but it is more expensive, therefore clustering can be used as a preprocessing method before choosing an attribute subset and classifying it. For instance, categorizing genes with comparable functioning or grouping customers based on their buying habits.

*Types of clustering methods:*

* Partitioning Methods
* Hierarchical Agglomerative (divisive) methods
* Density based methods
* Grid-based methods
* Model-based methods

## Predication

Prediction can be accomplished using the regression technique. The relationship between one or more independent variables and dependent variables can be modeled using regression analysis. Independent variables in data mining are characteristics that are previously known, whereas response variables are what we wish to forecast. Unfortunately, not all real-world issues can be predicted. Sales volumes, stock prices, and product failure rates, for instance, are all exceedingly challenging to forecast because they may be influenced by intricate interactions among numerous predictor factors. Therefore, it may be essential to estimate future values using more sophisticated techniques (such as logistic regression, decision trees, or neural networks). Frequently, the same model types can be applied to classification and regression. For instance, the decision tree algorithm CART (Classification and Regression Trees).

*Types of regression methods*

* Linear Regression
* Multivariate Linear Regression
* Nonlinear Regression
* Multivariate Nonlinear Regression

## Association rule

In order to uncover common item set finds across huge data sets, association and correlation are typically used. Business decisions like catalog design, cross-marketing, and customer shopping behavior research are all aided by these kinds of findings. Algorithms for association rules must be able to provide rules with confidence levels below one. Although there are a lot of potential Association Rules for a given dataset, the majority of them are often of little (if any) significance.

*Types of association rule*

* Multilevel association rule
* Multidimensional association rule
* Quantitative association rule

## Neural networks

Each connection in a neural network has a weight associated with it. It is made up of connected input/output units. The network adjusts its weights as it learns in order to anticipate the right class labels for the input tuples. Neural networks have the extraordinary capacity to extract patterns and identify trends from complex or ambiguous data that are too subtle to be seen by humans or other computer techniques. These are ideal for inputs and outputs with continuous values. For instance, rearranging handwritten characters, teaching a computer to comprehend English text, and many other real-world business issues have all been effectively implemented in numerous industries. Finding patterns is best done by neural networks.

*Types of neural networks*

* Back Propagation

## Data Mining Application

## Data mining is a very young technology that has not yet reached its full potential. Despite this, it is already utilized often in a variety of industries. Retail businesses, medical facilities, financial institutions, and insurance firms are a few of these entities. Many of these businesses combine data mining with other crucial techniques like statistics, pattern recognition, and others. Finding patterns and relationships that would be challenging to find without the use of data mining is possible. Many companies utilize this technology because it helps them understand their clients better and make wise marketing choices. Here is an overview of business issues and the data mining technology that was used to solve them.

## FBTO Dutch Insurance Company

Challenges

* To lower the price of direct mail.
* Improve the effectiveness of marketing initiatives.
* Increase cross-selling to current customers via inbound channels including the company's sales center and the internet, and conduct a one-year efficacy test of the solution.

Results

* The marketing staff was given the opportunity to forecast the success of its efforts.
* Enhanced the conception, optimization, and implementation of marketing campaigns.
* A 35% decrease in postage costs.
* A 40% increase in conversion rates.

## ECtel Ltd., Israel

Challenges

* Fraudulent behavior in the telecommunications industry

Results

 • Significantly decreased telecom fraud for more than 150 telecom firms globally.

 • By enabling real-time fraud detection, money was saved.

## Provident Financial’s Home credit Division, United Kingdom

Challenges

* No system to detect and prevent fraud.

Results

* Early fraud identification led to cost savings and decreased agent and consumer fraud frequency and severity.
* Reduced investigators' workload and raised the prosecution rate.

## Standard Life Mutual Financial Services Companies

Challenges

* Recognize the essential characteristics of customers drawn to their mortgage offer.
* Promote products from Standard Life Bank to customers of other Standard Life firms.
* Create a remortgage model that may be used on the group website to assess the profitability of the mortgage business that Standard Life Bank is accepting.

Results

* Created a propensity model that can be used across the whole group's prospect pool to identify important client types for the Standard Life Bank mortgage offer.
* Recognized the main factors that influence remortgage product purchases.
* Outperformed the control group's answer by a factor of nine using the model.
* Acquired $47 million worth of mortgage application revenue, or £33 million.

## Shenandoah Life insurance company United States.

Challenges

* The policy approval procedure was laborious and paper-based.
* There were delays in approval due to the distribution of these paper copies to different departments.

 Results

* Decreased the time needed to issue some policies by 20%.Provided management with up-to-date knowledge on policies that were pending.
* Process improvements for staff performance evaluations and under writing.

## Soft map Company Ltd., Tokyo

Challenges

* Customers had trouble selecting hardware and software, which was impeding online sales.

Results

* Following the launch of the recommendation engine, page visitors grew by 67% each month.
* Sales rose 18% during the same period the previous year, and profits more than tripled.

## Conclusion

In various corporate domains, data mining is important for identifying patterns, making predictions, discovering knowledge, etc. Classification, clustering, and other data mining techniques and algorithms aid in identifying patterns that can be used to predict future business trends. Data mining is regarded as one of the most significant frontiers in database and information systems and one of the most promising multidisciplinary advances in information technology since it has a broad application domain almost in every business where the data is generated.

## References

1. Jiawei Han and Micheline Kamber (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 Mining guide from [http://www.crisp-dm.org/CRISPWP-0800.pdf.](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.](http://www.spss.com/success/?source=homepage&hpzone=nav_bar)
5. [https://www.allbusiness.com/Techno](http://www.allbusiness.com/Technology/computer-software-data-management/633425-1.html)logy /com[puter-software-data-management/ 633425-1.html,](http://www.allbusiness.com/Technology/computer-software-data-management/633425-1.html) last retrieved on 15th Aug 2010.
6. [http://www.kdnuggets.com/.](http://www.kdnuggets.com/)