IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 5, Ver. I (Sep - Oct 2018), PP 61-65
www.iosrjournals.org

A Survey on Machine Learning And Data Mining Methods And Applications

Rajan Singh, Bramah Hazela
Department of Computer Science & Engineering, Amity University Uttar Pradesh
Corresponding Author: Rajan Singh

Abstract: *In This paper we have a tendency to describe a centered literature Survey of machine learning (ML) and data processing (DM) ways and applications .Under this analysis paper explanations of every machine learning and data processing technique square measure providing. It's established on the amount of the importance of a rising technique, Papers representing every technique were recognized, and it's additionally used for scan, and summarized .Because we know that the data are very important in ML & DM methods. As we know that ML&DM is very important and growing research area .it is used by the natural scientist to as well.in this paper we want explain and survey of combining both machine learning and data mining methods.*
Keywords*: Data mining, Machine learning ,Clustering.*

Date of Submission: 14-09-2018 Date of acceptance: 29-09-2018

I. Introduction

As we study about Data mining.it is extracting information and information from huge sum of data or warehouse. As we know that Data mining is an important step in searching knowledge from databases. There are many data bases, data marts, data warehouses in the world. If the data are not considered to find out the interesting patterns, then the data would become data graves. It is nearly difficult to extract the interesting unseen patterns in the sea of data without the help of data mining tools. As we know that there are seven steps use in data mining .They are data mining, data cleaning and data integration, pattern evolution, data selection, data transformation, knowledge presentation. Database technology had grown from original file processing to the improvement of data mining tools and their applications. Data mining are used in many field example business, management science and engineering ,government administration and environmental security .and many places, data mining and machine learning are very important area in research field of computer science .

Quick progress is due to the improvements in data analysis research, growing in the database industry and it is also very useful in the resulting market requirements for methods that are proficient of extracting valued knowledge from huge data stores. Machine learning is a developed and well-recognized research area in field of computer science, it is basically use for searching models, patterns, and symmetries in data. Machine learning can be approaches in two groups. First is symbolic and second are statistical approaches. Under symbolic approaches we study about decision tree and logical representation.in the case of Statistical approaches we study about k-nearest neighbor or Pattern-recognition methods, Bayesian classifiers neural network support vector machine.

As we know that machine learning is tree type of learning. Supervised learning and unsupervised learning and reinforcement learning. Under supervised learning we need to a guide but in the case of unsupervised learning we does not need a guide and in the case reinforcement learning we use agent and action and reward.

Important Area of data mining

1.1 Web mining

Web mining is very important area in data mining as we know that huge amount of data present on search engine. Data mining have a productive area for web mining. it is use for data mining methods for abstraction of info from web documents and services .we know that under web mining finding documents intentional for the Web .under web mining we do Selection and preprocessing of the information regained from the web. and we also did in web mining analysis.

1.2 Text mining

Text mining is additionally necessary space in data processing as we all know that text mining is a term of data mining or Knowledge Discovery in Text (KDT).The shapeless text may well be deep-mined victimisation data retrieval, text classification, or applying information processing techniques as a preprocessing step. Text Mining

involves several applications such text categorization, clustering, finding patterns and serial patterns in texts, linguistics, and association discovery.

1.3 Spatial Data mining

The abstraction data processing deals with knowledge associated with location. The explosion of geographically connected knowledge for fast development of IT, digital mapping remote sensing, GIS demands for developing databases for abstraction analysis and modeling. abstraction knowledge description, classification, association, clustering, trend, and outlier analysis of abstraction data processing.

1.4 Biological Data mining

As we all know that biological data processing is extremely helpful analysis space in data processing. Storage of clinical and biological information from DNA microarray information, genomic Sequences, macromolecule interactions still as sequences, electronic health records, sickness pathways, medical specialty pictures and also the list goes on .In the clinical context, biologists try to search out the biological processes that are the reason behind a sickness. There ar some problems associated with these high-dimensional biological information. These matters embrace screaky and incomplete information, integration numerous sources of information and process pc intensive tasks.

1.5 Educational Data mining

Under academic data processing we tend to cowl academic space of knowledge record. academic data processing (EDM) is associate degree rising analysis space involved with the distinctive styles of knowledge that come back from academic settings, and victimisation those strategies to higher perceive students. Academic data processing focuses on developing new tools and algorithms for locating knowledge patterns.

II. Machine Learning & Data Mining Techniques

2.1 Classification

Classification finds rules that partition information into some teams. The input for the classification is that the coaching set. The coaching set's category labels ar already legendary. Classification assigns category labels to unlabeled records supported a model that acquires information from the coaching datasets. Such categoryification is legendary is understood |is thought as supervised learning because the class labels ar known. There ar many classification models. a number of the common classification models ar call trees ,genetic algorithms and neural network, support vector machines (SVM), Bayesian classifiers. The Application includes credit risk analysis, fraud detection, banking and medical application.

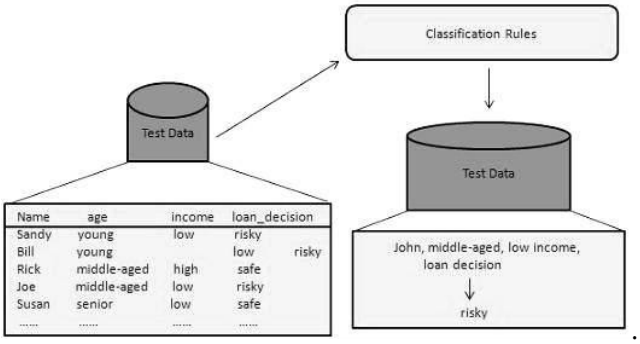
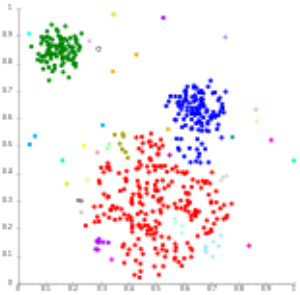


Fig.(1) - Classification of data

2.2 Clustering

As we know that Clustering is a method of grouping the information into cluster so things inside a cluster have almost similar. However arterribly dissimilar to the item in different cluster.disimilarities ar defendant supported the attribute values agglomeration the objects. Agglomeration algorithms is also used for organizing information, categorise information for model construction and information Compression, outlier detection, etc. several agglomeration algorithms were developed and ar classified as partitioning strategies, ranked strategies,density based mostly and grid based strategies.



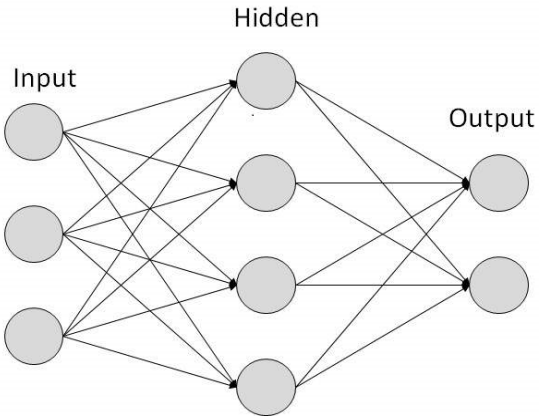
Fig(2)- Clustering of data

2.3 Association rule mining

Association rule mining realize attention-grabbing association or co-relation relationships among an oversized set of information things. With the large quantity information of knowledge of information unendingly being collected and hold on several industries area unit attention-grabbing in mining association rules from their data bases .the discovery of attention-grabbing association relationship among immense quantity of business group action records. They serving to in several business method process like catalogue style cross selling and drawing card analysis..

2.4 Neural network

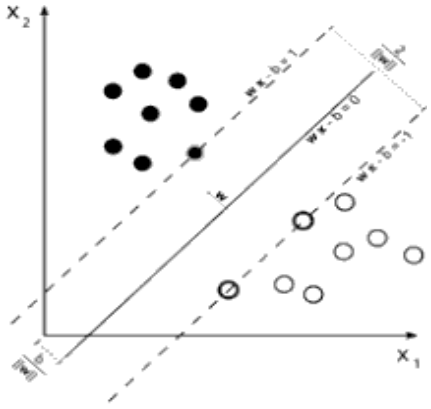
Neural networks area unit new computing Paradigm that's impressed by the biological system, like the brain, to method data. It involves developing mathematical structures with ability to find out. The Neural networks have the power to extract substantive and helpful patterns and trends from the advanced information. it's applicable to globe issues particularly just in case of business. because the neural networks area unit smart at distinguishing patterns or trends, they will be Applicable for prediction or foretelling desires. The system consists of extremely interconnected process components (neurons) operating along to resolve a particular drawback. Artificial neural network (ANN) learns by example. ANN is designed for specific application as classification, pattern recognition etc.



Fig(3)- Neural network

2.5 Support vector machine

Support vector machine can be used for non-linear problem. The goal of SVM is optimize sepreting hyper plane which maximize is margin of the training data. Under support vector machine we can take closet point of hyper plane. A sepreting hyper plane could be straight line that is used to seprate data in different classes. support vector machine can work with any no of daimension.in a given case it is possible to get several seprating hyper plan why maximize the distance of the data point from dicision surface the optimul hyper plan with the biggest margin. A margin is define as the distance of the closest point from the decision surface the support vector machine are the points that of closed decision no of closest vector.



Fig(4)-Support vector machine

2.6 Genetic algorithms

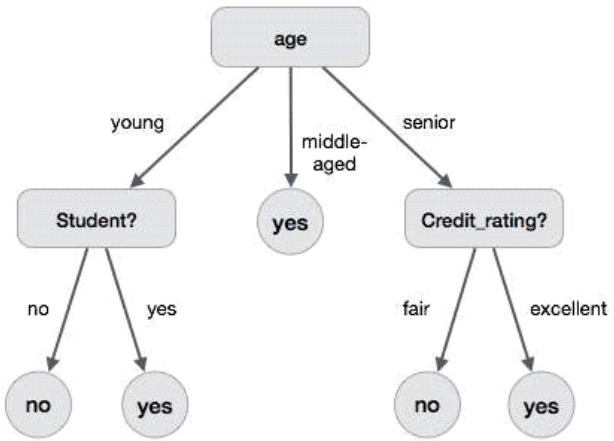
Genetic algorithms square measure a replacement paradigm in computing impressed by Darwin's theory of evolution. A population of the individual with doable resolution to a tangle is made at first arbitrarily. Then the crossover is finished by combining pairs of people to provide offspring of next generation. A mutation method is employed to change the genetic structure of some members of recent generation indiscriminately .The algorithmic rule searches for an answer within the serial generation. once Associate in Nursing optimum resolution is found or some fastened time is advance, the method involves a finish. Genetic algorithms square measure wide employed in issues wherever improvement is needed.

2.7 Naive Bayes classifier

As we know that this is a supervised classification methodology developed victimisation Bayes Theorem of contingent possibility with a 'Naive' hypothesis that each combine of article is reciprocally freelance. This is, in less complicated words, existence of a article isn't tormented by existence of additional by some suggests that. no matter this over-simplified hypothesis, Naïve Bayes classifiers performed great in several sensible things, like in text organization and spam Discovery. solely alittle quantity of coaching information is need to estimate bound limits. Beside, Naïve Bayes classifiershave significantly outdone even extremely innovative cataloging methods. .

2.8 Decision tree

In the decision tree .it is a classification tree famously called decision tree is one in all the first triumphant regulated learning equation. It makes a chart or tree that administrations fanning techniques to decide every plausible aftereffects of a decision . in an extremely call tests a component ,eachbranch compare to upshot of the parent hub and each leaf to end with doles out the classification mark.to arrange relate event ,a best down strategy is connected start at the premise of this tree .for an unmistakable element of hub , the branch understanding to the value of the information reason for that property is considered until the point when a leaf is contacted or a mark is set.



Fig(5)-Decision tree of data

III. Conclusion

In this paper we have a tendency to regarding examine all some aspects about machine learning and data processing .under this paper we have a tendency to describe the literature review regarding machine learning and data processing and their ways and application .mining is in a position to produce added answers and results. With reference to data processing analysis, per annum the analysis community addresses new open issues and new drawback areas, for several of that data processing is in a position to produce added answers and results. As a result of the entomb disciplinary nature of knowledge mining, there's an enormous flow of latest data, widening the spectrum of issues which will be resolved by the employment of this technology. Another reason why data processing contains a scientific and business future was given by Friedman (1998): "Every time the quantity of knowledge will increase by an element of ten, we should all rethink however we have a tendency to analyze it." To achieve its full business exploitation, data processing remains lacking the standardization to the degree of, as an example, the standardization offered for database systems. There square measure initiatives during this direction, which is able to diminish the monopoly of high-priced closed-architecture systems. For data processing to be actually successful it's necessary that data processing tools become offered in major info product further as in normal desktop applications. alternative necessary recent developments square measure open supply data processing services, tools for on-line construction of knowledge mining workflows, further because the language and ingredients of knowledge mining through the event of a knowledge mining In the future, we have a tendency to imagine intensive development and exaggerated usage of knowledge mining in specific domain areas, like bioinformatics, multimedia, text and internet data analysis. On the opposite hand, as datamining may be used for building police work systems, recent analysis conjointly concentrates on developing algorithms for mining databases while not compromising sensitive info. A shift towards automatic use of knowledge mining in sensible systems is additionally expectedto become quite common.

References

[1]. Ayodele, T. O. (2010). Types of machine learning algorithms. In *New advances in machine learning*. InTech.

[2]. Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Yang, B., Betteridge, J. & Krishnamurthy, J. (2018). Never-ending learning. *Communications of the ACM*, *61*(5), 103-115.

[3]. 3Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.

[4]. Wang, J., Jebara, T., & Chang, S. F. (2013). Semi-supervised learning using greedy max-cut. *Journal of Machine Learning Research*, *14*(Mar), 771-800.

[5]. Chapelle, O., Sindhwani, V., & Keerthi, S. S. (2008). Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research*, *9*(Feb), 203-233.

[6]. Baxter, J. (2000). A model of inductive bias learning. *Journal of Artificial Intelligence Research*, *12*, 149-198.

[7]. Wu, Q., Ding, G., Xu, Y., Feng, S., Du, Z., Wang, J., & Long, K. (2014). Cognitive internet of things: a new paradigm beyond connection. *IEEE Internet of Things Journal*, *1*(2), 129-143.

[8]. Chen, X., Shrivastava, A., & Gupta, A. (2013). Neil: Extracting visual knowledge from web data. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1409-1416.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

* Rajan Singh,. " A Survey on Machine Learning And Data Mining Methods And Applications." IOSR Journal of Computer Engineering (IOSR-JCE) 20.5 (2018): 61-65.