A REVIEW ON MACHINE LEARNING IN DATA MINING

Sreelakshmi K P[1], Sayana Anil Kumar[2] and Vanitha T[3]

Abstract--Data has grown so fast and is still growing in an exponential phase that a serious measure has to be taken to see to it that its value is utilised to the fullest. It is a very clear fact that, a promising technology is crucial in maintaining and manipulating this data. Here we focus on Data mining with Machine learning and the various fields where it is applied.

Keywords: Data mining, Machine learning, dataset, analysis.

I.INTRODUCTION

Data generation is unstoppable and ever growing in this modern era of Big Data. Today high priority is given to data and its value which could be referred to as the treasure of 21st century. Big data is one of the most promising paradigm in science where, with a right tool one can unlock the right information out of it, but to carry out this task, it is not practical to analyse and obtain the data out of a massive dataset simultaneously in complex systems like social media database, search engines, and public service systems etc. This is where Machine learning is used in Data mining. Here the trained set of artificial nodes can imitate the task of trained neurons present in the brain of a human being. The machines are trained or allowed to be self-taught with a set of training data till the machine becomes capable of carrying out the task without any errors.Data mining is used to collect the important details and information from a large set of data which is similar to finding concealed data from a storehouse. Data Mining is not restricted to any particular field or dataset. Likewise machine learning could also be implemented where real time data mining is inevitable.

II MACHINE LEARNING

Machine learning systems automatically learn programme from data. [9]. Machine learning is used in web search, airports, biomedical science etc. Machine learning can be classified into Supervised learning, Unsupervised learning, semi-supervised learning, reinforcement learning and transduction learning based on the way the algorithms are trained to carry out a task. It covers a vast area research in the fields of artificial intelligence, statistics,

[1] *Department of Software Technology Aloysius Institute of Management and Information technology, Mangalore, Karnataka , India.*
[2] *Department of Software Technology Aloysius Institute of Management and Information technology, Mangalore, Karnataka , India.*
[3] *Department of Software Technology Aloysius Institute of Management and Information technology, Mangalore, Karnataka , India.*

neurobiology as well. In a technologically advanced world where data is getting piled up there is a high demand that is being placed to process and mine data at a faster phase, with high accuracy and complete automation. This is where machine learning techniques are being adopted and made use to obtain a better results out of random data. This paper focus on machine learning being used in the field of Data mining.

A. Heuristic model to improve feature selection based on machine learning in data mining

Jahinet al., in the year 2016,discussed machine learning based on three major factors , that are Experience 'E', task 'T' and performance 'P'. Machine learning can be classified into two. One is the supervised learning in which input algorithm provide a procedure to forecast the result of input whereas the other one is the unsupervised learning in which the concealed design is identified. There are many problems in traditional data handling methods such as property selection, in such cases machine learning is utilized in Data mining. Machine learning algorithms have been used to improve data classifications and pattern recognition in data mining especially for feature selection[1].

B. Data mining with machine learning applied for email deception

Sujeetet al., in the year 2013, discussed that machine learning in data mining is applied for email deception. From past few years junk emails has become major problem for the sustainability of internet. For this a number of spam detection algorithms have been proposed, but they were not up to the mark. With machine learning algorithms the data analysis is carried out by classifying spam words as an identification mark. These patterns are unique and close patterns which are extracted from the message content and also it is shown that by this feature there is a good impact in classifying the spam from legitimate emails. This research also proves that the method can be easily implemented with respect to popular algorithms like Logistic Regression, Neural Network and so on. There are many positive outcomes which shows higher accuracy by using Machine learning compared to related methods, a method has been proposed where the analysis is mainly focused on the body of the email containing spam words to target the users and it is evaluated using linguistic analysis tool. The pre-processing and the classification is the important step followed by this method. Overall, the algorithm perform better that predicted as it provide better results. Future work on this can also be extended by introduction of a learning system within the collaborative filter[2].

C. A syllabus on Data Mining and Machine Learning with Applications to cyber security

Epishikinaet al., in the year 2016, discussed about a syllabus for a new university course in the area of data mining and machine learning with applications to cybersecurity. Here it is mainly focused on providing students with fundamental concepts in data mining (including neural networks and support vector machines etc. currently, development of Intelligent security (ISS) is mainstream in the field of cybersecurity [3] . Machine learning for creation and usage of ISS, deep understanding of mathematical and logical foundations of Big Data analytics is required. The cyber security and machine learning when combined into one single concept has lot of potential in providing serious protection to the people. Today scientist combine these two fields and several serious research are being carried out in this field throughout the world.

D. Mining Process Control Data Using Machine Learning

_____

Nasret al., in the year 2009, discussed about Mining Process Control Data Using Machine Learning. Manufacturing industry databases usually consist of process control, process procedure and quality control data[4].Data Mining can provide aggressive advantage in the manufacturing field because traditional technique that solve the process control problems are very time consuming. Therefore data mining technique is used to deal with process control problem easily. For classification of data, machine learning technique is used commonly. It consist of two stages namely learning stage and testing stage. In learning stage, a model(idea) is built by an algorithm called learner which can be utilized in the second stage. In testing stage, based on the model of the first stage they try to divide new specimen. In this way the machine learning is implemented in data handling.

E.   Transduction and typicalness for quality assessment of individual classifications in Machine Learning and Data Mining

Kukaret al., in the year 2004, discussed about quality evaluation of individual classification. Usually machine learning algorithm output only bare predictions for the new unclassified examples[5]. There is no method to provide evaluation for classification, but machine learning algorithm provide this feature. The focus is on the entire performance of the classifier because it is very helpful in financial, critical control application etc. The machine learning algorithm uses a different approach by collecting result from variety of classifiers. In the machine learning technique the probability distribution for decision making by assigning probability 1 to predicted class and probability 0 to remaining classes is being used.

F.   Datamining with machine learning for intelligent airport system:

Ariyawansaet al., in the year 2016,discussed about machine learning for airport system.. As airport creates a lifelong impression of its people. But behind the scene there are so many complex computational tasks that are being processed and analysed for the maintenance of accurate and highly responsible system in the world. To overcome the challenges, airports implemented isolated solutions, but they fail frequently. This section focus on researches that was carried out by putting all the entries in one platform which is dependent on Data mining which predicts the fight delay, passenger profile etc by analysing the data[6].

G.   *Crawler intelligence with machine learning and data mining integration*--

Abhirajet al., in the year 2015, discussed about crawler intelligence. Data mining is an inevitable task in the working of search engines. Better indexing of webpages and the contents present in it makes a search engine user-friendly and popular. With the enormous amount of information present in the internet, data mining and indexing is not something that could be carried out manually, in-fact it is impossible to imagine such a task being carried out manually. This is where the Crawler Intelligence comes into picture. The traditional methods of indexing webpages were just based on sophisticated algorithms that managed the whole process of indexing the content of web by real-time analysis of dataset. Due to the need for more accuracy and perfection, the concept of Data mining with machine learning is gaining wider importance these days. Crawling being a data intensive activity, data mining algorithms play an important role in introducing intelligence in crawlers[7].

With the widening of application that manipulates the data present in the internet, there is an increase in the pace at which the demand for the development of intelligence based tools is

growing. This is exactly where the data mining with the help of machine role gains its importance. Data can be heterogeneous and contains higher value within it, this makes it hard for the development of machine learning algorithm that can manipulate this heterogeneous data. One of the challenges faced by machine learning is the need for a higher accuracy by reducing the errors caused by the machines.

H.   *Data Mining using a machine learning library in c++*
Kohaviet al., in the year 1997, discuss about solutions to analyse the heterogeneous data with a new approach where Machine learning library class in C++ is used to identify and understand the algorithm that can provide higher efficiency in analysing the provided dataset. This approach is very crucial in terms of huge heterogeneous data when it is collected from multiple sources and must be analysed and processed simultaneously.
One of the major advantage of MLC++ is that it makes the job of developers easy to develop algorithms that can be highly specific and specially intended to carry out a particular task (dataset analysis). Developers can use MLC- to create new algorithms suitable for their specific task[8]. There is a tremendous importance to specialized algorithms because there are no algorithms that can be categorised as to be able to perform all the task in a complex environment where the data is rich in variety.

III. CONCLUSION

With the volume of Information growing larger each day, where traditional methods of data analysis are getting replacedby complex and intelligent algorithms that can handle the data, machine learning is gaining much prominence in Data mining. Estimating the current situation and the amount of work being carried out in this area, it is clear that the future is governed by brilliant algorithms that can think and make correct decisions based on experience, the applications of machine learning in various fields of data mining is an evidence to this. This shows that there is a never ending scope for Machine learning in an ever growing world of Information.

REFERENCES

[1]   Majumdar, Jahin, Anwesha Mal, and Shruti Gupta. "Heuristic model to improve Feature Selection based on Machine Learning in Data Mining." *Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference*. IEEE, 2016.
[2]   More, Sujeet, and S. A. Kulkarni. "Data mining with machine learning applied for email deception." *Optical Imaging Sensor and Security (ICOSS), 2013 International Conference on*. IEEE, 2013.
[3]   Epishkina, Anna, and Sergey Zapechnikov. "A syllabus on data mining and machine learning with applications to cybersecurity." *Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), 2016 Third International Conference on*. IEEE, 2016.
[4]   Nasr, Emad S. Abouel, and Hisham Al-Mubaid. "Mining process control data using machine learning." *Computers & Industrial Engineering, 2009. CIE 2009. International Conference on*. IEEE, 2009.
[5]   Kukar, Matjaz. "Transduction and typicalness for quality assessment of individual classifications in machine learning and data mining." *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*. IEEE, 2004.
[6]   Ariyawansa, Chamath Malinda, and AchalaChathurangaAponso. "Review on state of art data mining and machine learning techniques for intelligent Airport systems." *2016 2nd International Conference on Information Management (ICIM)*. IEEE, 2016.
[7]   Darshakar, Abhiraj. "Crawler intelligence with Machine Learning and Data Mining integration." *Pervasive Computing (ICPC), 2015 International Conference on*. IEEE, 2015.
[8]   Kohavi, Ron, Dan Sommerfield, and James Dougherty. "Data Mining Using a Machine Learning Library in C++." *International Journal on Artificial Intelligence Tools* 6.04 (1997): 537-566.
[9]   Domingos, Pedro. "A few useful things to know about machine learning." *Communications of the ACM* 55.10 (2012): 78-87.