Original Article

Mining And Machine Learning In The Manufacturing Products

Madhavi Kolukuluri[1], A Sagar[2], Sureddi Niharika[3], Chepala kishore[4], Nakka Krishna Prasad[5]

[1]Proffessor, Dept. of Computer Science & Engineering (Data Science), Nadimpalli Satyanarayana Raju Institute ofTechnology, Sontyam, Visakhapatnam-531173, Andhra Pradesh, India.
[2,3,4,5] Student, Dept. of Computer Science & Engineering (Data Science), Nadimpalli Satyanarayana Raju Institute ofTechnology, Sontyam, Visakhapatnam-531173, Andhra Pradesh, India.
DOI: 10.47750/pnr.2022.13.S07.635

Abstract

Today, we live in a data-driven world where millions of knowledge resources come from the Internet and various databases. Manufacturing companies need to use different types of techniques and tools to achieve their foundational goals. In this context, the use of machine learning (ML) and data mining (DM) techniques and tools could be very helpful to overcome manufacturing challenges. This paper presents a comprehensive literature review on how machine learning techniques can be appliedto implement manufacturing mechanisms. Our contributions are intended to provide an understanding of the main approaches and algorithms that have been used to improve manufacturing processes in recent decades. In addition, the main steps of the Knowledge Discovery in Databases (KDD) process to be followed in manufacturing applications are explained in detail.

Keywords-- Foundational goals, data mining, comprehensive literature, knowledge discovery.

I. INTRODUCTION

Machine learning (ML) is an important research area in artificial intelligence that helps computers build models based on experience and accurately predict future events. The main ML approaches fall into two main categories: supervised learning and unsupervised learning. A typical problem of supervised learning is classification, while unsupervised learning is more commonly used for clustering problems. Commonly used classification techniques include neural networks, support vector machines, and decision trees, and the most commonly usedclustering technique is k-means. ML Techniques have been successfully applied in many different fields such as health, education, wireless sensor networks, and finance. Thispaper provides an overview of the use of ML techniques inmanufacturing [12].

In modern manufacturing plants, powerful data acquisitionsystems are used to electronically capture and transmit datafrom almost every process in the enterprise. Many production variables are measured continuously at variousstages, and their values are stored in company databases. These data may relate to the characteristics of the products,the machines, the production line (e.g., which machine wasused with which setting parameters), the personnel operating the production line (e.g., the experience level ofthe worker, the type of shift), the raw materials used in theprocess, the environment (humidity,

temperature, etc.), the sensors attached to the machines (vibration, force, pressure, voltage, etc.), machine failures/maintenance, product quality, and other importantproduction factors.

Recently, several reports have appeared on ML and DM inthe manufacturing industry. However, some of them focusonly on one area of manufacturing, such as the electronicsindustry, additive manufacturing, or semiconductor manufacturing. Some of them focus only on one topic, such as quality assessment; some of them include only data mining studies but no machine learning; some of them arenot

comprehensive; and some of them are older. Unlike previous studies that have the same goal, this article provides a systematic overview of machine learning and data mining in manufacturing, covers the current state of research, includes a comprehensive list of available studiesin this field, clearly states the benefits and challenges in manufacturing, and opens new perspectives for future applications.

II. DATA MINING

Data Mining (DM) is defined as the process of analyzing largedatabases, usually data warehouses or internet, to discover newinformation, hidden patterns and behaviours. It's an process ofanalyzing huge amounts of data to discover hidden traits, patterns and to predict future trends for the business applications [1]. Data Mining analyse datasets of rational databases, in multiple dimensions and angles, producing asummary of the general trends found in the dataset, and demand for the products. Mining information and knowledge from largedatabases has been recognized by many researchers as a key research topic in industrial companies as an important area withan opportunity of major revenues [3]. Enterprises must make business critical decisions based on large datasets stored in their databases and directly affect decision-making. By using past data stored in databases we built machine learning models to predict the demand for the particular product, So we can increase the production of that product.

III. WEB MINING

The other mining technique is Web Mining. Web Mining is defined as extraction of relevant information and hidden patterns found in World Wide Web. Web mining is actually anarea of data mining related to the information available on internet. Users prefer World Wide Web more to download andupload data [4]. Over the last few years, we have witnessed anexplosive growth in the information available on the Web. Webmining could be viewed as the use of data mining techniques toretrieve, extract, generalize, and analyze data to discover hidden traits, patterns and to predict future trends for the business applications [1]. Different tools and algorithms are used for extraction of data from web that includes web documents, images etc. Web mining is rapidly becoming very important due to size of text documents increasing over the internet and finding relevant patterns, knowledge and informative data is very hard and time consuming if it is done manually [4]. Some processes included in web mining are:

Information Retrieval is a process of retrieving relevant and useful information over the web. Information retrieval has morefocuses on selection of relevant data from large collection of database and discovering new knowledge from large quantity of data to response user query. Information Retrieval steps includes searching, filtering and matching [4].

Information extraction is an automatic process of extracting analyzed data (structured). Information extraction is a task thatwork same like information retrieval but more focuses onextracting relevant facts [4].

Machine Learning is support process that helps in mining datafrom web. Machine learning can improve the web search by knowing user behavior (interest). Different machine learning methods are used in search engine to provide intelligent web service. It is much more efficient than traditional approach thatis information retrieval. It is a process that has ability to learn user behavior and enhance the performance on specific task.

Web mining is mainly classified into three types:

- Web Content Mining
- Web Structure Mining
- Web Usage Mining

Web Content Mining is a process of the knowledge extraction from the web documents or their descriptions available on the internet. The search for the knowledge on the Internet is difficult and time consuming task. This is what Web Content Mining is dealing with. It

is based on a combination of featuresof information retrieval, machine learning, and data mining.

Web Structure Mining is a process of finding structural information on the Internet. This type of mining examines the relationship between web pages based on the connections between them.

The constructed models can be used to categorize and find similar web resources, and for detecting copyright sites.

Web Usage Mining is the automatic detection of patterns in thepath of the user movement and the associated data collected oracquired as a result of interaction with one or more web sites. This direction is based on the extraction of data from the Web server logs. The purpose of the analysis is to identify thepreferences of the visitors using those or other resources on theInternet.

(i) Web Content Mining

Content Mining is a process of Web Mining in which needful informative data is extracted from web sites (WWW). Contentincludes audio, video, text documents, hyperlinks and structured record. Web contents are designed to deliver data tousers in the form of text, list, images, videos and tables. Over last few decades the amount of web pages (HTML) increases tobillions and still continues to grow. Searching query into billions of web documents is very difficult and time-consumingtask, content mining extracts queried data by performing different mining techniques and narrow down the search data which become easy to find required user data [4].

(ii) Web Structure Mining

Now a day's massive amount of data is increasing on web. World Wide Web is one of the most loved resources for information retrieval. Web mining techniques are very useful to discover knowledgeable data from web. Structure mining isone of the core techniques of web mining which deals with hyperlinks structure. Structure mining basically shows the structured summary of the website. It identifies relationship between linked web pages of websites. Continues growth of data over the internet become a challenging task to find

Web Mining Categories	**Techniques**	**Tools**	**Algorithms**
Web Content Mining	- Unstructured Data Mining - Structured Data Mining - Semi – Structure Data Mining - Multimedia Data Mining	- Screen Scaper - Mozenda - Automation Anywhere7 - Web Content Extractor - Web Info Extractor - Rapid Miner	- Decision Tree - Naive Bayes - Support Vector Machine - Neural Network
Web Structure Mining	- Link-based Classification - Link-based Cluster Analysis - Link Type - Link Strength - Link Cardinality	- Google PR Checker - Link Viewer	- Page Rank Algorithm - HITS algorithms (Hyperlink Induced Topic Search) - Weighted Page Rank Algorithm - Distance Rank Algorithm - Weighted Page Content Rank Algorithm - Webpage Ranking Using Link Attributes - Eigen Rumor Algorithm - Time Rank Algorithm -Tag Rank Algorithm - Query Dependent Ranking Algorithm

Web Usage ining	- **Data Preprocessing** Data Cleaning User & Session Identification - **Pattern Discovery** Statistical Analysis Association Rules Clustering Classification Sequential Patterns - **Pattern Analysis** Knowledge Query Mechanism OLAP (Online Analytical processing) Intelligent Agents	- **Data Preprocessing Tools** • Data Preparator • Sumatra TT • Lisp Miner • Speed Tracer - **Pattern Discovery Tools** • SEWEBAR-CMS • i-Miner • Argunaut • MiDas(Mining Internet Data for Associative Sequences) - **Pattern Analysis Tools** • Webalizer • Naviz • WebViz • WebMiner • Stratdyn	- **Association Rules** • Apriori Algorithm • Maxi-mal Forward References • Markov Chains • FP Growth • Prefix Span - **Clustering** • Self-Organized Maps • Graph Partitioning • Ant Based Technique • K-means with Genetic algorithms • Fuzzy c-mean Algorithm - **Classification** • Decision Trees • Naïve Bayesian Classifiers • K-nearest Neighbor Classifiers • Support Vector Machine - **Sequential Patterns** • MIDAS (Mining Internet Data for Association Sequences) algorithm

informative and required data. Web mining is just a data miningwhich digs data from the web. Different algorithmic techniquesare used to discover data from web. Structure mining analyzeshyperlinks of the website to collect informative data and sort out in categories like similarities and relationship. Intra-page isa type of mining that is performed at document level and at hyperlink level mining is known as inter-page mining. Link analysis is an old but very useful method that is way its value increases in the research area of web mining – Structure analysis is also called as Link-mining [4].

(iii) Web usage mining

Web usage mining also called log mining is a process ofrecording user access data on the web and collect data in form of logs. After visiting any website user leaves some informationbehind such as visiting time, IP address, visited pages etc. Thisinformation is collected, analyzed and store in logs. Which helps to understand user behavior and later can improves website structure. Web usage mining is a technique that automatically archives access patterns of user and thisinformation is mostly provided by web servers which are later collected in access logs. Logs stores much needed informationlike URL address, visiting time, Internet Protocol addresses etc.which can help an organization to understand their customer's behavior and insure good

service quality. Web usage mining dig and analyze data presentin log files which contains user access patterns. Main purpose of web usage mining is to observer user behavior at the time ofhis interacting with web. There are two types of pattern tracking
i.e. general tracking and customized tracking. In general tracking information is collected from web page history. In customized tracking the information is gathered for specific user [4].

IV. MACHINE LEARNING

The major role of machine learning starts primarily with Data mining and Analysis. The results of this analysis can form the basis of a report or a machine learning model, but it all starts with data. Machine Learning works by identifying the relationships between the data values that describe the characteristics of its features and the labels we want to predict and encapsulating these relationships in a model through a learning process

Machine Learning works by identifying the relationships between the data values that describe the characteristics of its features and the labels we want to predict and encapsulating these relationships in a model through a learning process.

Usually, a machine learning task is designed to establishinformation about a particular scenario or to test a hypothesis. For example, suppose a college professor collects data from machine learning students, including the number of courses taken, hours spent studying, and final grade on the end of termexam.

The professor could then take a sample of the data and analyze it to determine if there is a relationship between the amount of study a student undertakes and the final grade they obtain. They can use the data to test the hypothesis that only students who study for a minimum number of hours can expectto achieve a passing grade or even prepare the data to train a machine learning model that predicts the performance grade ofa student according to his/her study habits.

So, in machine learning, we have to train algorithms to find patterns and relationships between massive amounts of data tomake decisions and predictions based on new data. The better the algorithm performs, the more accurate decisions and predictions will become as it processes more data. This is whatthe role of Machine Learning is.

V. WEB CRAWLING

Web crawlers are mainly used to create a copy of all the visitedpages for later processing by a search engine that will index thedownloaded pages to provide fast searches. Crawlers can also be used for automating maintenance tasks on a Website, such as checking links or validating HTML code. Also, crawlers canbe used to gather specific types of information from Web pages,such as harvesting email addresses. A Web crawler is one typeof bot, or software agent. In general, it starts with a list of URLsto visit, called the seeds. As the

crawler visits these URLs, it identifies all the hyperlinks in the page and adds them to the listof URLs to visit, called the crawl frontier. URLs from the frontier are recursively visited according to a set of policies. There are three important characteristics of the Web that makecrawling it very difficult: • its large volume, • its fast rate of change, and • dynamic page generation, which combine to produce a wide variety of possible crawlable URLs. The large volume implies that the crawler can only download a fraction of the Web pages within a given time, so it needs to prioritize its downloads. The high rate of change implies that by the timethe crawler is downloading the last pages from a site, it is verylikely that new pages have been added to the site, or that pageshave already been updated or even deleted. As Edwards et al. noted, "Given that the bandwidth for conducting crawls is neither infinite nor free, it is becoming essential to crawl the Web in not only a scalable, but efficient way, if some reasonable measure of quality or freshness is to be maintained.". A crawler must carefully choose at each step which pages to visit next. The behavior of a Web crawler is theoutcome of a combination of policies: • A selection policy thatstates which pages to download.

• A re-visit policy that states when to check for changes to thepages.

• A politeness policy that states how to avoid overloadingWebsites.

• A parallelization policy that states how to coordinate distributed Web crawlers.

The importance of a page for a crawler can also be expressed as a function of the similarity of a page to a given query. Web crawlers that attempt to download pages that are similar to eachother are called focused crawler or topical crawlers. The main problem in focused crawling is that in the context of a Web crawler, we would like to be able to predict the similarity of thetext of a given page to the query before actually downloading the page[6].

VI. CONCLUSION:

Web data mining is a new technique to know the hidden patterns, marketing strategies and increase the profit margin. This method is the efficient way to prevent the losses to the company. The major advantage of this method is it is very easyto implement. on basis of this we can develop the business strategy according to it.

VII. REFERENCES:

[1] Abdul-Aziz Rashid Al-Azmi, (2013) "DATA, TEXT, ANDWEB MINING FOR BUSINESS INTELLIGENCE: A SURVEY" Department of Computer Engineering, Kuwait University, Kuwait.

[2] Yuefeng Li, Ning Zhong, "Web Mining Model and Its Applications for Information Gathering" Department of Systems and Information Engineering, Maebashi Institute of Technology, Maebashi 371-0816, Japan.

[3] Ming-Syan Chen[1], Jiawei Han[2], Philip S. Yu[3], "Data Mining: An Overview from Database Perspective" Elect. Eng.

Department National Taiwan Univ. Taipei, Taiwan, ROC[1],School of Computing Sci. Simon Fraser University B.C. V5A 1S6, Canada[2], IBM T.J. Watson Res. Ctr. P.O.Box 704 Yorktown, NY 10598, U.S.A[3].

[4] Muhammd Jawad Hamid Mughal, "Data Mining: Web Data Mining Techniques, Tools and Algorithms: An Overview" Department of Computer Science SZABIST DubaiCampus Dubai, United Arab Emirates.

[5] V. Bharanipriya[1], V. Kamakshi Prasad[2], "WEB CONTENT MINING TOOLS: A COMPARATIVE STUDY"

[6] Aman Kharwal, "AN INTRODUCTION TO MACHINELEARNING" January1,2021.

[7] Web Site : http://ww.celi.it

[8] Soumen Chakrabarti, "Mining the Web : Discovering Knowledge from Hypertext Data, Morgan Kaufmann,2003

[9] Bing Liu, "Web Data Mining: Exploring Hyperlinks, Contents and Usage Data", Springer, 2007

[10] Hsincun Chen and Michael Chau, “Web Mining : MachineLearning for Web Application” , Annual Review of Information Science and Technology, University of Arizona
[11] Chakrabarti, S. (2000). Data mining for hypertext: Atutorial survey. SIGKDD Explorations, 1(1), 1-11

[12] AlicanDogan[1], DeryaBirant[2], “Machine learning and data mining in manufacturing” Dokuz Eylul University, The Graduate School of Natural and Applied Sciences, Izmir, Turkey[1], Dokuz Eylul University, Department of Computer Engineering, Izmir, Turkey[2]