

Exploration on the Algorithms in Data Mining using AI Techniques

M.Swathy¹, K.Subash², D.S.Ravi³

^{1*} Dept.Computer Science, St. Joseph's College, Bharathidasan university, Trichy, India.

² Dept.Computer Science, St. Joseph's College, Bharathidasan University, Trichy, India.

¹ Dept.Computer Science, St. Joseph's College, Bharathidasan university, Trichy, India.

*Corresponding Author: mswathymohan94@gmail.com, ph.: 7397170154

Available online at: www.ijcseonline.org

Abstract— Artificial Intelligence(AI) is an emerging technology which brings innovation to the society by simulating human intelligence which are processed by machines. This AI technique is classified into two types such as Supervised and Unsupervised where the supervised classify the data using networks and an unsupervised uses genetic algorithm for the finding of hidden relationship between data. The data mining, which is used in artificial intelligence for various purposes such as segmentation, classification of data, diagnosis of images and also prediction problem. This paper gives a detailed description about the survey of machine learning techniques in artificial intelligence by using remote sensing imaginary mining and pattern recognition technique.

Keywords- Artificial Intelligence, machines, supervised, unsupervised, mining

I. INTRODUCTION

a) Artificial Intelligence Definition

The founder of artificial intelligence is John McCarthy [1]. AI is a mechanism which used for the purpose of science for making the machines more intelligent and also especially making the computer program more intelligent. It is also a way of making intelligent computer, knowledge controlled robots, and making the software to think intelligently. AI is ways of making the machine to think and act like humans in many ways. The goals of AI are to create an expert system for behaving and work like the human being. The various agents for artificial intelligence are human agent, robot agent and software agent.

The categories of AI are conventional AI and computational intelligence. The conventional AI is also used in statistical methods used to implement the techniques.

B) Artificial Intelligence Feature

The main features for the AI are representation and reasoning about knowledge. Ability to communicate from one to another [2]. Huge search spaces for the genetic algorithm. Reasoning with conflicting the data, ability to learn and adapt the situation.

c) Types of Intelligence

Based on the different areas, the intelligence is classified into six types[3]. Those are linguistic intelligence is the ability to speak and orate things, logical –mathematical intelligence is the ability to understand, the ability for the objects, action hidden in data and musical intelligence explains about the communication of rhythm spatial

intelligence explains about third and fourth factors, bodily kinesthetic intelligence explains about the body moments and inter personal and intra personal intelligence explains about the own feeling and the other feeling.

d) Data Mining In Artificial Intelligence

Data mining [4] is a technique which provides tools for new innovations, for visualizing, pattern recognition, for the data which helps in decision making and also identifies the trend in the data. They are many techniques in data mining such as association rules, Bayesian technique, Decision rules, and Gaussian model. These are the various techniques which are used to solve classification and segmentation issues. In order to inspect the huge amount of data present in the organization two techniques are followed those are supervised and unsupervised techniques. The supervised technique deals about the different data which is given and the unsupervised deals about the genetic algorithm where the hidden data are classified. There are various algorithms for data mining technique those are follows knowledge discovery where the various knowledge are discussed related to data. Decision tree algorithm explains about the various trees branching algorithm, the next one is Bayesian theorem where it consists of posterior probability and priori probability.

II. REVIEW OF LITERATURE

a) Navale et al. [5] has proposed that the stock market prediction was one of the most difficult tasks in those days. Nowadays, applications are accepted through online for the exchange of shares, the next step will be for predicting the

values of each share. Their algorithm explains about the prediction of shares which are stored in the database. Prediction is done through ARMA (Autoregressive-moving-average) algorithm. Though the methods in data mining and artificial intelligence are effective in accuracy the performance parameter has to be increased.

b) Nelson Sizwe et al. [6] has proposed that knowledge is gathered only by the collection of data which is difficult. Many organizations are collecting huge number of data and converting it to information in real time. For extracting such data warehousing is more important and to extract the needed information mining technique is needed. Their techniques which are implemented here are knowledge representation, acquisition and inferences. The interpretation of vast amounts of data using domain is lacking.

c) ArunaChamatkar et al. [7] has proposed that the goal of data mining is to extract the new information from the database. Any technique has to be biased for the approach. This mining technique consists of patterns, structure and singularities for the growing amount of data. The tool which is used, here are data mining that is derived from an AI algorithm which helps to convert the data in specified manner such as structure and patterns. Though the data is structured architecture for each network has to do carefully.

d) Xindong Wu et al. [8] have proposed that knowledge discovery and data mining are interrelated to each other not only in data, but also in statistics and machine learning. Their paper explains about analysis of large volumes of data in the database. Fraud detection has to be controlled.

e) Katarina Hilovska et al. [9] have proposed that AI has vast area of interdisciplinary area such as medicine, transportation. Their paper explains various types of tasks stored in this technique by using fuzzy system, ANN, Genetic Algorithm. Though data the parameters are in high level it is useful only for short time of period.

f) Pinky Solanki et al. [10] has proposed work on categorization of images using similarity present in each image. Though it is easy for humans to categorize it is difficult to machine so for that purpose Nearest Neighbor algorithm and means clustering algorithm are used. Though is accuracy is satisfactory the algorithm consumes more time.

g) Haoyan Hong et al. [11] has proposed work on the map for suspecting of floods in Poyang country using data mining methods. This is used to identify non flood area and the cause for flooding. The technique which is used in their paper is Fuzzy Wolf –SVM model for evidence. Though it has best accuracy the natural cause offloaded is not notified.

h) MuneoKushimo et al. [12] has proposed work on the level of care for the patients who are in need for long term health care. The level is divided into five parts where level one includes vocabulary such as recreation, toilet,

morning and afternoon. Level five consists of tube, danger, treatment, removal and discovery. Though the treatment is classified by various levels. This was some of the major problem are not handled.

i) Maurice Dawson et al. [13] has proposed work on internet enabled technology, which helps the terrorist to increase the terrorism in Africa and also to find out who is behind this dreadfulness. The reason behind this terrorist attack is with the help of data which is collected by using (OSINT). This is done with the help of the R language and python language. This only deals with software, but they are other ways for attacking such as political parties and various universities.

j) Jorespires et al. [14] has proposed work on the information which is gathered and organized for the cognitive learning is done by schemata technique. Their algorithm main aim is to provide a full learning using cognitive learning by genetic algorithm. It is not implemented in other platform and blind and deaf people can't use this.

k) Elyanerubio et al. [15] has proposed work on the decrease of autopsies in the modern world. The technique which is implied here are association rule in data mining. The real data has to be collected.

l) PhillipeBondel et al. [16] has proposed how to compare the morphological data of a submarine using domain terrain models and also to identify using patterns and trends. Though remote sensing plays a vital role in this accuracy was still need to improve.

m) Ricardo vascarda et al. [17] has proposed that the safe and requiring of skin temperature using medical infrared thermal images using Bayesian algorithm. Though the images are clear the view of sample image is also needed in this technique.

n) Sushmarao et al. [18] has proposed the admission for the students using the geographical area and the admission year. Their algorithm explains about the Bayesian rule which classifies about year of students who are going, but the admission percentage is not counted in this technique.

o) MainazFaridi et al. [19] has proposed the spatial data mining technique for the agricultural intelligence for the purpose of increasing of human population and decrease of crop. The technique which is used is a spatial association of data mining in geographical information system to identify the water surrounded area. The accuracy of the water land has to be identified.

I. Comparison of various Data Mining Algorithms, Techniques and Limitations

Author	Methods	Limitation
Navale [5]	Prediction of the stock market using ARMA	Using ARMA the prediction is high.
Nelson size we[6]	Knowledge acquisition using artificial intelligence	Using warehousing and mining technique the data are analyzed

Arunachamatkar[7]	Mining of data using AI technique.	Patters, structures and singularities are used.
Xindongwu[8]	Knowledge discovery using data mining in the data base.	Statistics and machine learning are used.
Katarina hiluska[9]	Categorization of algorithm[5]	Nearest neighbour algorithm
Rinkysolanki[10]	Suspecting floods using fuzzy wolf - svm algorithm.	Fuzzy wolf- svm algorithm
Haoyoanhong[11]	Level of care using data mining.	Using a Bayesian algorithm the level is determined
Muneokushima[12]	OSINT in artificial intelligence.	Open software's are used for the purpose of obtaining the various details.
Maurice Dawson[13]	Cognitive technique in genetic algorithm.	Schemata technique is used for analyzing the data
Jorespires[14]	Atopsies using data mining technique.	Association technique is used.
Elyanepires[15]	Submarine mode using a terrain model.	Domain terrain model is used for submarine purpose.
Phillipebondel[16]	Medical infrared thermal images.	Using baysianalgorithim the images are scanned.
Ricardo vascarda[17]	A student admission technique using a Bayesian algorithm.	Using mining technique the rate of admission is counted
Sushmarao[18]	Data mining for agricultural intelligence.	Spatial data mining is used

IV.CONCLUSION

Large volume of data from knowledge discovery is a great area in AI and data mining. The sustained area of research by various algorithms. Prediction and optimization are done in genetic algorithm. These tools are used only for short period of time. The most important technology in AI which leads for decision making fuzzy systems is used for the insurance. In this paper, the Bayesian algorithm is used for pattern recognition, this would help both researchers and academicians to know more about data mining. Many researchers are also proposed this algorithm in their work. Though, the pattern recognition is still critical problem for more researchers can be done to enhance this area.

REFERENCES

- [1] "Artificial Intelligence Tutorial" https://www.tutorialspoint.com/artificial_intelligence/index.htm
- [2] "Artificial Intelligence" https://en.wikipedia.org/wiki/Artificial_intelligence
- [3] "Artificial intelligence" <http://searchcio.techtarget.com/definition/AI>
- [4] "Artificial intelligence" <https://www.edx.org/micromasters/columbiacx-artificial-intelligence>
- [5] G.S.Navale,NishanthDudhwala,KunalJadhav,PawanGabda, brijkishorevihangam, "prediction of stock market using data mining & AI" *International Journal of Computer Applications* (0975 – 8887) Volume 134 – No.12, January 2016
- [6] Nelson.Sizwe,Madonsela,paulin.Mbecke,CharlesMbohwa., "Integrating Artificial Intelligence into Data Warehousing and Data Mining" *Proceedings of the World Congress on Engineering and Computer Science* 2015 Vol II WCECS 2015, October 21-23, 2015.
- [7] Ms. Aruna J. Chamatkar " An Artificial Intelligence for Data Mining" *IOSR Journal of Computer Science (IOSR-JCE)* e-ISSN: 2278-0661, p-ISSN: 2278-8727 PP 53-57
- [8] Xindong Wu1, "Data Mining: An AI Perspective" *IEEE Computational Intelligence Bulletin* December 2004 Vol.4 No.2
- [9] KatarinaHilovska "Application of Artificial Intelligence and Data Mining Techniques to Financial Markets" *ACTA VSFS*, 2012, vol. 6, issue 1, 62-76.
- [10] PinkiSolanki and GirdharGopal "Image Categorization Using Improved Data Mining Technique" *Big Data Analytics, Advances in Intelligent Systems and Computing* 654, https://doi.org/10.1007/978-981-10-6620-7_19
- [11] Haoyuan Hong ParaskevasTsangaratosIoanna Ilia ,JunzhiLiu,A-Xing Zhu, Wei Chen "Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China" *Science of the Total Environment* 625 (2018) 575–588(2017)
- [12] MuneoKushima, Kenji Araki, Tomoyoshi Yamazaki, Sanae Araki, TaisukeOgawa,NoboruSonehara "Text Data Mining of Care Life Log by the Level of Care Required Using KeyGraph" *Proceedings of the International MultiConference of Engineers and Computer Scientists* 2017 Vol I, IMECS 2017, March 15 - 17, 2017
- [13] Maurice Dawson, Max Lieble, and AdewaleAdeboje "Open Source Intelligence: Performing DataMining and Link Analysis to Track TerroristActivities" *Information Technology – New Generations, Advances in Intelligent Systems and Computing* 558, DOI 10.1007/978-3-319-54978-1_22
- [14] Jorge Pires, Manuel Pérez Cota, Álvaro Rocha and Ramiro Gonçalves "Towards a New Approach of Learning:Learn by Thinking Extendingthe Paradigm Through Cognitive Learning and Artificial Intelligence Methods to Improve Special EducationNeeds" *Developments and Advances in IntelligentSystems and Applications, Studies in Computational Intelligence* 718, DOI 10.1007/978-3-319-58965-7_18
- [15] Elayne Rubio Delgado, Lisbeth Rodríguez-Mazahua,Silvestre Gustavo Peláez-Camarena, José Antonio PaletGuzmánand AsdrúbalLópez-Chau "Association Analysis of Medical Opinions About the Non-realization of Autopsies in a Mexican Hospital "New Perspectives on Applied Industrial Tools and Techniques, Management and Industrial Engineering,DOI 10.1007/978-3-319-56871-3_12
- [16] Philippe Blondel "Quantitative Analyses of Morphological Data " *Submarine Geomorphology, Springer Geology*" DOI 10.1007/978-3-319-57852-1_5
- [17] H.S. SushmaRao, Aishwarya Suresh and VinayakHegde "Academic Dashboard—Descriptive Analytical Approach to Analyze Student Admission Using Education Data Mining " *Information and Communication Technology for Sustainable Development, Lecture Notes in Networks and Systems* 10, https://doi.org/10.1007/978-981-10-3920-1_43
- [18] MainazFaridi, SeemaVerma and Saurabh Mukherjee "Integration of GIS, Spatial Data Mining, and Fuzzy Logic for Agricultural Intelligence" *Soft Computing: Theories and Applications, Advances in Intelligent Systems and Computing* 583,https://doi.org/10.1007/978-981-10-5687-1_16