

Artificial Intelligence in Data Mining & Big Data

Salem Alaisawi
Salem.k.m@gmail.com

ABSTRACT:

Data mining and big data could be a new and chop-chop growing field. It attracts ideas and resources multiple disciplines, together with machine learning, statistics, information analysis, high performance computing and commerce. This explains the dynamic, many-sided and chop-chop evolving nature of the data mining discipline. whereas there's a broad accord that the abstract goal of information mining is to discover new and helpful data in knowledge bases this can be wherever the accord ends and the suggests that of achieving this goal are as diver seas the communities contributive. The foundations of all data processing methods, however, are in arithmetic. Any moderately sized treatment of information mining techniques

necessarily needs to be selective and perhaps biased towards a specific approach. data processing techniques are wont to notice patterns, structure or regularities and singularities in massive and growing datasets. Artificial neural network ANN are gross simplification of real networks of neurons. The paradigm of neural network that Data mining and big data could be a new and chop-chop growing field. It attracts ideas and resources multiple disciplines, together with machine learning, statistics, information analysis, high performance computing and commerce. This explains the dynamic, many-sided and chop-chop evolving nature of the data mining discipline. whereas there's a broad accord that the abstract goal of information mining is to discover new and helpful data in knowledge bases this can be wherever the accord ends and the suggests that of achieving this goal are as diver seas the communities contributive. The foundations of all data processing methods, however, are in arithmetic. Any moderately sized treatment of information mining techniques necessarily needs to be selective and perhaps biased towards a specific approach. data processing techniques are wont to notice patterns, structure or regularities and singularities in massive and

growing datasets. Artificial neural network ANN are gross simplification of real networks of

neurons. The paradigm of neural network that began throughout the 1940's guarantees to moderately sized treatment of information mining techniques necessarily needs to be selective and perhaps biased towards a specific approach. Data processing techniques are wontto notice patterns, structure or regularities and singularities in massive and growing datasets. Artificial neural network ANN are gross simplification of real networks of neurons.

INTRODUCTION:

Artificial intelligence in data mining and big data techniques are wide utilized in several domain to resolve classification, planning, prediction, optimization problems, diagnosis, computation, collecting and analyzing customer information, gleaning insights into what customers want and need, and acting on those insights. The aim of this special issue is to reflect the most recent development during this analysis field and supply advanced information for researchers actively functioning on algorithms and applications of artificial intelligence. So, data mining and big data is present almost everywhere and it is very important and essential to preserve the data that is generated in huge amount so something should not be missed out. Artificial intelligence is often used to process this type of data. Basically, Artificial Intelligence and its sub branches (For example Machine Learning, Deep Learning, Neutral Networks), all are algorithm based. These algorithmic methods are used on huge amount of Data (Big Data) to produce desired results and to find trends, patterns and predictions. Complex analytical tasks faster than human imagination are done on Big Data with the help of AI. (Baker and Inventado 2014) artificial neural network developed with a scientific stepwise procedure that optimizes a criterion usually referred to as the learning rule. The input/output training information is key for

these networks because it conveys the data that is important to get the optimum in operation purpose. Additionally, a nonlinear nature creates neural network process parts a versatile system. One representative definition is pivoted round the comparison of intelligence of computing machines with people in general. Another definition is anxious with the performance of machines that is traditionally are judged to lie among the domain of intelligence. So, we can consider artificial intelligence in data mining and bigdata. Its application is to method knowledge and appraise them. That covers any input data, that implicitly needs some structure, to perform some algorithm. The rule will split into ordered or parallel and perform over the computer process unit or the graphics processing unit. the selection on which of them effects the performance of the rule in speed. As any rule, the terminal result's Associate in multiple output. A changed definition could be a terminal result of multiple output. The performance is entirely mechanical and programmed into the computer. it's the interpretation by human of the information in context, the rule method, and the results that make value. Associate in analysis is open concluded as report, analysis, or prediction. It depends on the expertise of the person who performs a data mining. At a really high level, AI revolves around an artificially created entity performing or thinking in an autonomous manner like humans. Big Data has to do with parsing and analyzing large data sets to look for trends, patterns, etc. they're only related in so far as you can use one technology to compliment the other. For example, instead of humans deciding how to interpret, refine and act on big data analysis, AI could be used instead to make these decisions. Conversely, Big Data could be used by AI in its self-learning and/or decision making. (Turban, Sharda et al. 2010)

BASIC CONCEPTS:

Real-world information discovery processes sometimes comprise of one or a lot of knowledge pre-processing, machine learning, evaluation, and visual steps. Hence,

a data mining and bigdata platform ought to permit complicated experiment styles, clear knowledge handling, comfortable parameter handling and optimization, versatile rearrangements, and extendibility. Finally, a speedy prototyping environment for information discovery should ensure that even complicated experiments will simply be designed. During this section we are going to discuss a number of the fundamental ideas of Yale that eased the look of recent strategies(Mobasher 2007).

DESCRIPTION:

Artificial Intelligence, in the present, are complex and effective but nowhere near human intelligence. Humans use the data present around them, and the data accumulated in the past to figure out anything and everything. However, AIs don't have that ability just yet. AIs just huge data dumps to clear their objectives. This means that AIs require a huge pool of data to do something as simple as editing letters.

Data Science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms. This means that data science helps AIs figure out solutions to problems by linking similar data for future use. Fundamentally, data science allows for AIs to find appropriate and meaningful information from those huge pools faster and more efficiently.

An example of this is Face book facial recognition system which, over time, gathers a lot of data about existing users and applies the same techniques for facial recognition with new users. Another example is Google's self-driving cars which gathers data from its surroundings in real time and processes that information to make intelligent decisions on the road.

Machine Learning is likely the connection between data science and artificial intelligence since machine learning is the process of learning from data over time. However, it's not the only thing connecting those two together. But machine learning is the branch of AI that works best with data science. And big data is present almost everywhere and it is very important and essential

to preserve the data that is generated in huge amount so something should not be missed out. It is utmost difficult to store that massive amount of data any company generates. Traditional computing techniques are not able to handle such large datasets. Artificial intelligence is often used to process this type of data. Basically, Artificial Intelligence and its sub branches (For example Machine Learning, Deep Learning, Neural Networks), all are algorithm based. These algorithmic methods are used on huge amount of Data (Big Data) to produce desired results and to find trends, patterns and predictions. Complex analytical tasks faster than human imagination are done on Big Data with the help of ML and AI.

In the past, AI wasn't developed because of small data sets. There wasn't any real life or real time data. Today, as massive amount of data sets is always available, this has derived towards AI. Nowadays, Artificial Intelligence and Big Data have formed a relationship. You can say, there is no Artificial Intelligence without Big Data.

here's real-time, always-available access to the data and tools that enable rapid analysis. This has propelled AI and machine learning and allowed the transition to a data-first approach.

almost people think AI is about complexity of algorithms, "Huge mistake" it's basically about lot of filtrated and revised data. *AI* is a way to navigate and gather insights in the world of *Big Data* (Purohit, Banerjee et al. 2015)

CLASSIFICATION AND PREDICTION:

Classification is that the method of coming up with a group of models to predict the category of objects whose class label is unknown. The derived model could also be pictured in varied forms, like if-then rules, call trees, or mathematical formulas.

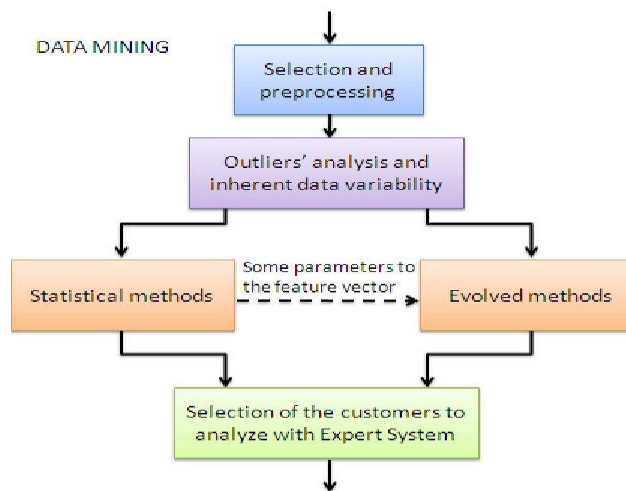
A decision tree could be a flow-chart-like tree structure wherever every node denotes a look at on associate degree attribute worth, every branch represents associate degree outcome of the look at, and every tree leaf represents a category or

class distribution. call trees is regenerate to classification rules.

Classification is used for predicting the category label of information objects. Prediction encompasses the identification of distribution trends supported the obtainable knowledge.

The data mining method consists of associate degree repetitive sequence of the subsequent steps. The data mining method consists of associate degree reiterative sequence of the subsequent steps:

1. information coherence and cleansing to get rid of noise and inconsistent data.
2. information integration specified multiple data sources is also combined.
3. information choice wherever data relevant to the analysis are retrieved.
4. information transformation wherever data are consolidated into forms applicable for mining.
5. Pattern recognition and applied math techniques are applied to extract patterns.
6. Pattern analysis to spot fascinating patterns representing data.
7. image techniques are wont to gift well-mined data to users. (Srinivas, Rani et al. 2010)



LIMITS OF DATA MINING AND BIG DATA:

GIGO (garbage in garbage out) is nearly continually documented with relation to data processing, because the quality of the data gained through data processing relies on the standard of the historical information. we all know information inconsistencies and managing multiple data sources represent massive issues in data management. Information improvement Another limitation of data mining is that it solely extracts data restricted to the precise set of historical knowledge, and answers will solely be obtained and taken with regards to previous trends learned from the info. This limit one's ability to learn from new trends. as a result of the choice tree is trained specifically on the historical knowledge set, it doesn't account for personalization inside the tree. to boot, data processing (decision trees, rules, clusters) are non-incremental and don't adapt whereas in production. (Provost and Fawcett 2013)

CONCLSION

The field of specific applications are aimed to extract specific information. The domain experts by considering the user's necessities and different context parameters guide the system. The intelligent interfaces

and intelligent agents up to some extent build the applying generic however have limitations. The domain consulta decisions at completely different stages are influenced by the factors like field and knowledge details, aim of the data mining, and the context parameters. Ann offers a robust and distributed computing design, with important learning talents and they are ready to represent extremely nonlinear and multivariable relationships. However, ANN are helpful particularly once there's no a priori information regarding the analyzed knowledge. Efforts in artificial intelligence resulting in valuable new services and applications (Kitchin 2014)

REFERENCES:

- Baker, R. S. and P. S. Inventado (2014). Educational data mining and learning analytics. Learning analytics, Springer: 61-75.
- Hall, M., et al. (2009). "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter **11**(1): 10-18.
- Kitchin, R. (2014). "Big Data, new epistemologies and paradigm shifts." Big data & society **1**(1): 2053951714528481.
- Mobasher, B. (2007). Data mining for web personalization. The adaptive web, Springer: 90-135.
- Provost, F. and T. Fawcett (2013). "Data science and its relationship to big data and data-driven decision making." Big data **1**(1): 51-59.
- Purohit, H., et al. (2015). "Gender-based violence in 140 characters or fewer: A# BigData case study of Twitter." arXiv preprint arXiv:1503.02086.
- Srinivas, K., et al. (2010). "Applications of data mining techniques in healthcare and prediction of heart attacks." International

Journal on Computer Science and Engineering (IJCSE) 2(02): 250-255.

Turban, E., et al. (2010). "Decision Support and Business Intelligence Systems (required)." Google Scholar.