Quantum Computing In Big Data

RASHMI RAJ

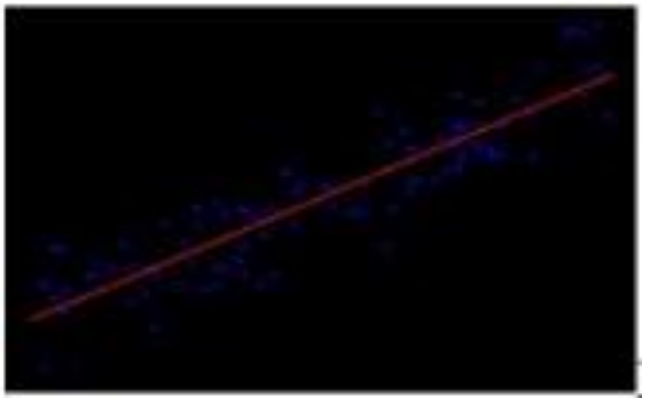
ABSTRACTION: We live in a world which celebrates data intrusion at all levels. We like to see data generated and imbibed. We also need order where efficiency in data handling is desired. So big data was envisaged to make data look relevant and be managed. The extent of data created is large. We would not be able to handle such a amount of data if we do not create placeholders. For placeholders to work we need better data encoding. For encoders to go we need level adept handlers. Quantum computing steps in. It gives enormous data handlers and it makes big data look simpler. When we have quantum coding to the necessary we live in world where we are able to achieve data sufficiency.

INTRODUCTION TO BIG DATA

We live in a data fed world. Thousand of data is generated to be analysed by industry. Unless we don't analyse data it becomes obsolete. It becomes dark data. Dark data is of no use. It is data which needs processing. It depends on the requirement and future trend for data to be properly analysed. Big data is a new field in computing where massive data generated is to be analysed for monetisation. We get analysis done by data processing. When we talk of big data we think in terms to petabyte and tera bytes. The magnanimity of data makes data analysis a very herculean task. We cant help go data waste because each data tells us something about its successor. Also every data give a clue to customer behaviour. Useless we don't analyse customer behaviour we wont be able to leverage advantage. So it becomes imperative for the company to gain competitive advantage over others by analysis of big dat. Big data has three sources. First one being the data generated by user itself. When he logs on to system, he generates data which is called user related data. Secondly when he uses internet of things like mobile and utility devices then too data is generated. When interaction with web generates data, it is called internet of thing data. Third is the data aggregated when by third party resources. That data is also crucial because it reflects users mindset. When we talk of positioning competitive advantage we think of data as tool to predict customers net usage. Data has certain parameters like value velocity, veracity, variety and volume. Each one points to data manipulation. Velocity is term used to measure rate of data growth,.. value refers to data as a money. Veracity gives clue to data differentiation. Volume makes us understand the sheer magnanimity of data. Data in itself can never be read in isolation. Data needs to stored and should be manipulated by algorithms for data to be useful for public. For that to happen we further need four parameters. They are segment, accuracy, dependably and lastly its authenticity. By segment e mean clustering data in classes and using a single variable to getting a regressive value. By regression we mean isolating one variable to ascertain complex behaviour of class. Then also we need to check class dependency or class veracity. If data is organised in a class then it has to be authentic. The dependably and authenticity gives a edge over competitors. For manipulation of data we also make use of algorithms. We use basic algorithms like regression and cluster methods. Lage companies employ pretty complex techniques for data mining. Walmart was one of large companies which started using data mining. When hurricane stuck America then the buying trend of Americans was checked. They realised that a certain product was favoured among all. Then Walmart started selling those product and it gain massive profit when hurricane strike next. In this way data is manipulated and data mining is done which proves to useful in long run. Quantum comping is very aggressive field. Its value as a computing can be guessed from the fact that it processes petabytes of data in a fraction seconds. As we are advancing we are moving towards a society which needs data for leads. We have to understand the fact data as such can't make lot difference unless its analysed properly. Complex computing results is todays reality. When we are doing weather reporting or dealing with stock or talking about electricity grid then massive data is generated. That's big data. Quantum computing comes to rescue. It's not possible for conventional computing to get a big picture of what disparate data actually mean. Modern computing relies heavily on quantum computing. It has to make data monetised by applying algorithms meant for discrete computing. Quantum computing depend on discrete mathematics. The basic unit of data storage in quantum computing is qubit. The states are superimposable and we get a compounded state set which gives the n bit qubit variable. Quantum computing is very fast growing segment of todays modern computing because it gets its state by curvature of photon. Quantum computing can solve big data problem. We can get faster results in nanoseconds and where time variable matters quantum computing does a great job. In business where every piece data of is important. Quantum computing steps into give a complete view of what data can do. Some open source software like qisbit can deal with quantum computing. As it is growing field many researches are invited to get a valid data report.

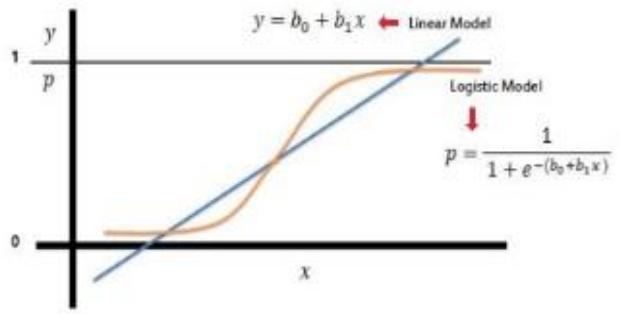
Algorithms

Algorithms are used to process data. We use random algorithm to make data look familiar to us. When we derive work from data we use algorithm to make it look plausible. Big data has to do with massive data level. The amount of data generated in big data is so large that to derive any work from it require optimal use of algorithms. Regression steps in to gain meaning out of disparate data. By regression we generally mean to get a relationship between two sets of independent variable and dependent variables. The two data sets can be infinite and finite depending on the problem set. When we are establishing a relationship between the two then we derive mathematical function which determines how the two data variable can change or a change in one variable set will reflect on other. Regression technique is optimally used in forecasting and prediction. We are talking about regression we generally mean a parameterised function and its observational data. The casualty and effect relation is more often talked about.



Model showing regression points.

Regression techniques has of two types. One linear regression and the other one is logistic regression. Both differ in two way. The difference is established in the observational data set and the way a function responds to the change in variable. Whereas linear regression deal with infinite data when we speak of logistic data it has to with finite sets of data. Linear regression is quite often used in business models but in big data analytics we frequently use logistic regression. linear regression is has its variables as a scalar variable set. As the name suggests we have to deal with a linear model in establishing relation or look a the linear cause and effect between two sets of data. Whereas logistic model gives regression analysis which is more of continues variability, linear regression give a scalar point of view. Linear regression is also used ins situations when a error reduction is required. When we have continuously rectify and get a pattern which is a very simplistic view of data relation we use linear regression. Similarly logistic regression has to deal with another level. We take binomial relationship or multinomial relational ship in this model. it was developed by David cox. The above function shows a typical linear regression function. While it represents regressive coefficients the summation of errors are given by the sigma. Coefficients are given by x1 and x2 and if we include more independent variables. The graphical analysis of both regression models gives a clear picture. While linear regression gives straight line on y variable the logistic regression produces the sigmoid graph. The quality of graph in more the models differs.



Algorithms used in big data analytics: k-means clustering algorithms

Clustering algorithms as the word suggests work with data clustering. The data clusters are made such that each cluster shows one variable which has to be mapped to one variable and get the its closets mean distance so that we make veronoi cells. By veronoi cells we mean the generation of cells which work on distance form mean principal. It finds its use in machine learning and weather forecasting. One of the very interesting feature is vector quantisation. When are using k clustering algorithm we should consider segment. As the word goes we have discrete segments based on variable k. Closer samples where we produce colour pallet is based on variable k and other good uses like image segmentation and computer graphics are also good ways to suggest that k clustering algorithm is a sure way to segregate data in a cluster. Clustering depends on variable k which is hard to find. Also is the fact that this algorithm is used in supervised and unsupervised learning where every data cluster has to be revisited to get learning done. Euclidean distance and the variable k defines cluster algorithm . Euclidian distance is a form of straight distance between two points in Euclidean space. It is of the metrics used for k clustering segregation.

Association rule mining algorithm

Association rule mining algorithm is a good way to judge attributes based on associations on data similarity in data structure. Data is mapped in such away so as to derive meaning insight which give boost to industry. To get a insightful understanding of data specially in business mean we use association rule mining algorithm. In this algorithm we establish rule between two sets of data groups. This has its way in business atmospheres. When we talk of relations we try to find interesting meanings to data being given. It helps in forecasting and it also helps in bioinformatics and data mining. It has its application in other industries too. Here we look for relations and if relation is established a amount of confidence is generated. If this confidence is more then we move to level of conviction. Getting a good probability is conviction while finding a relation is support. So these four concepts of support, confidence, conviction and lift rules. A minimum support is found between datasets which gives a degree of confidence. Like if we say a man is sure to buy Potatoes and tomatoes from supermarket then through probability function we try to say that he is likely to buy hamburger and cheese. Here through probability function we bolster our claim.

Support vector machine

Support vector machine is a special way to segregate data depending on hyperplane. There exists a superficial hyperplanes which divide the data into its distance measured from mean. Superficial hyperplane is in a divide between two data groups. We can have hyperplane which will segregate data . Support vector machine is for supervised learning. In machine learning this algorithm is often used. Machine learning requires constant categorisation of data. From a set of points we derive a (p-1) set of hyperplane which determines how a data has to be utilised. when we have a pair of data set depending on kernel function, its segregation is decided. We often use this support vector algorithm in character imaging and handwriting analysis. We are using its potential to determine image recognition. When we have massive data points then their virtual separation is almost impossible we have the concept of soft margin classifier. It's a simple way to classify data based on their distance from k variable of kernel function. Maximal margin classifier is that line which determines the efficacy of data separation.

Quantum computing

In twentieth century data has become new jackpot. Whosoever has data is sitting on gold mine. In this century we saw many classical theories of physics being bended. Nowadays we talk of quantum computing when we mean massive data handling. Classical physics work on binary stares. With mushrooming of data especially in business, state became plural as context also grew simultaneously. Quantum computing defines more states by superimposition where every state pair gives way to many others states pairs just by superimposition. photon which is a unit of light is defined it in vector state. Vector space is time space graphs . Polarisation of light produces many vector states because light photons can move in any degree. Trigonometric functions and calculus functions are needed for every state. In compounding state we can derive the present state by using probability function and algorithms. The basic unit of data in quantum computing is qubits. It can be arrived by mathematical function. There can be of any variable length, especially in cases where is no compromise on memory. Quantum computing is done in petaflops time scale and it give result with accuracy.

Mathematics behind quantum computing

Qubit - a qubit is a basic unit of data storage. it is actually a device which gives different states of data in a orthogonal vector space. The states are given by two pronged variables. the are a0/0> + a1/1>. A0 and a1 are complex numbers. /a0/2 + /a1/2. The two states are given by probability of /ao /and a1 while a0 and a1 are lost. The state is determined by a gate. How the information flows is determined by the gate. There are various gates like one qubit gate and two qubit gate. One qubit gate- one qubit gate is ate where the output vector b0 and b1 is determined by initial state. There final state must be being consonance with initial state. The values are given below. The not gate is the simplest from of gate that does something.

INPUT OUTPUT
/0> /1>
/1> /0>
($a_0b_0a_0b_1$
$a_1b_0a_1b_1$).
Two qubit gate is something different.

INPUT OUTPUT
/0>
$\frac{1}{\sqrt{2}}$/0> +
$\frac{1}{\sqrt{2}}$/1>
/1>
$\frac{1}{\sqrt{2}}$/0>-
$\frac{1}{\sqrt{2}}$/0>
/0>
/0> = (10
0 0),
/0>
/1> = (01
0 0),
/1>
/0> = (0 0
1 0),
/1>
/1> = (0 0 0 1)

Quantum computing in big data

In todays world data is generated at a faster speed. We generate exabytes of data every single day. It means we generate data in accordance with 90 minutes of films being run. To handle this vast amount of data new theories are needed. Big data is very challenging. It is so humongous that traditional computing or classical physics cannot do justice. We need novel solution to get right answer. Thus quantum computer steps in. Moors law suggest that microprocessor speed should double very 18 years .To justify amount of data being produced and to give an edge to Moore's law quantum computing is the solution. Speed is needed to solve problems at a larger level. Because we need answers to every problem at a very faster rate, size and complexity of datasets is also increasing. To accommodate large data sets quantum computing is needed. Big data changed the we look at data. Quantum computing too is doing the same. The easy way to look at the data is organizing it in a way that result is got in fraction of second. Lot of data interaction occurs and million o data is produced daily as users interact with computer. So that data has to be well organised. Traditionally we had to rely on transistors for production of data but today quantum fabrication has made data look different. By entanglement we mean paradox. Any property of quantum particle will have its paradox to. If one particle has a spin then there will exists a correlated reverse property too. The application duality of wave particle is also important. The Heisenberg theory of duality of nature too comes into picture. At the quantum level both can co exists. Qubit the quintessential data particle gives a clue to n bit variable existence. Today quantum computing is present in Toshiba and ID quntique. After miniaturisation of transistors, this is the next steps in data sophistication. There is a bridge between classical theory and quantum computing that generates electron spin explaining Heisenberg's theory.

CONCLUSION

The Data war goes on. The quest for better data management is on. Not only we need data to be managed but investment in newer technology too is required to fulfill quest. Data as a resource is a good concept. When we get data pool we actually think it as a resource or an asset or a precursor for change. As a token for technological progress we must go for data handlers . Not just data handlers at a gross level but we need better data handlers at quantum level too.

ACKNOWLEDGEMENT

I express my gratitude towards my almamator GURU GHASIDAS UNIVERSITY, which gave me values and courage to visualize this paper. My thanks also goes to my family and friends whose constant support provided me enough nurture. My thanks goes to my special friend Nita without whom this paper would not have been possible

REFERENCE

[1] Montanaro Ashley Quantum algorithms: an overview npj Quantum Information volume 2, Article number: 15023 (2016)

[2] R. Akis and D. K. Ferry Quantum waveguide array generator for performing Fourier transforms: Alternate route to quantum computing Appl. Phys. Lett. 79, 2823 (2001)

[3] TY - JOUR
AU - Teh, Hui Yie
AU - Kempa-Liehr, Andreas W.
AU - Wang, Kevin I-Kai
PY - 2020
DA - 2020/02/11
TI - Sensor data quality: a systematic review
JO - Journal of Big Data
SP - 11
VL - 7
IS - 1
AB - Sensor data quality plays a vital role in Internet of Things (IoT) applications as they are rendered useless if the data quality is bad
SN - 2196-1115
UR - https://doi.org/10.1186/s40537-020-0285-1
DO - 10.1186/s40537-020-0285-1
ID - Teh2020
ER -

AUTHOR PROFILE

I got my MS degree from GURU GHASIDAS UNIVERSITY BILASPER in 1999. I have been giving guest lectures in DISCRETE MATHEMATICS and COMPUTATIONAL LOGIC in various places.