International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume 4, Issue 3 (March '2018)

THE PLAY OF MATHEMATICS IN BIG DATA ANALYTICS

K.L. Subhiksha Narayanan,
II-Bsc Mathematics,
Sri Krishna Adithya College of Arts and Science,
Coimbatore, Tamilnadu, India.

Abstract: The availability of huge amounts of data is often considered as the fourth industrial revolution we are living right now. And data accumulation allows us to tackle a wide range of social, economic, industrial and scientific challenges. But extracting meaningful information from the available data is not a trivial task. Well mathematics plays an important role in the existing algorithms for data processing through the techniques from mathematics. Here, I review some of the important applications of mathematics in big data analytics.

Keywords-Graph Theory, Recurrence Relation, Topology-persistence homology, matlab.

I.INTRODUCTION

In recent years Big Data has become more than just a buzz word. Every field produce huge amount of data. This data, has to be generated, acquired, stored, then repeatedly processed and analysed. But the thing is people started facing lots of problems in storing this giant amount of data called big data and this is becoming a serious issue in data storage. well, these problems can be solved with mathematics. The basic algorithms are been created with the help of mathematics. Mathematical tools are an important enabler in big data settings Development in big data settings not only require more computer speed and memory capacity, but also new advanced mathematics.

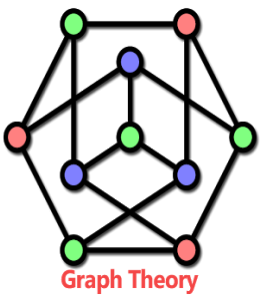
What is big data?

Big data is a new area which covers many disciplines and technologies. It is a data set that are so voluminous and complex that the traditional data processing application software are inadequate to deal with them.

Play of math:

Graph theory:

Data structure is a data collection for organising data in computer memory and these structure includes array, list, stack, tree, graph..



Graph is also a one like data structure representing relationship between connected objectsG=(V,E) is the basic formula of graph theory. In which graph G is an ordered pair of a set V of vertices and a set E of edges. The graph theory plays an important part in big data well, the half part of big data is covered with the social networking site and this works with the concept of graph theory.

For example: talking about the common word of today, the Facebook which covers the huge amount of big data which works with the graphical structure. And this kind of graphical structure is called as undirected graph.

Tree structure:

A tree is a special kind of graph. It is one of the basic structure in storing data. And it is also a powerful tool for organising data objects. In computer science, tree is a widely used abstract data type (ADT).

Programmatically we can define node in c or c++ as:

```
struct Node {
    int data;
    Node* left;
    Node* right;
};
```

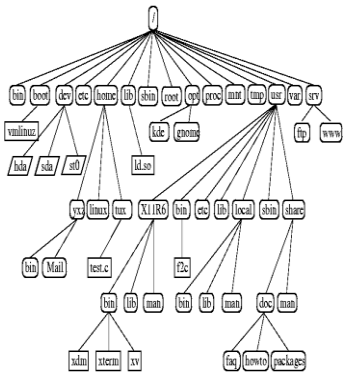
Tree data structure used in many ways and it is mainly used to manipulate hierarchical data it makes information easy to search. It manipulates sorted list of data.

Spectral graph theory is mainly used in imagining and data clustering, were the data clustering is much more important in storage of data which the big data is all about.

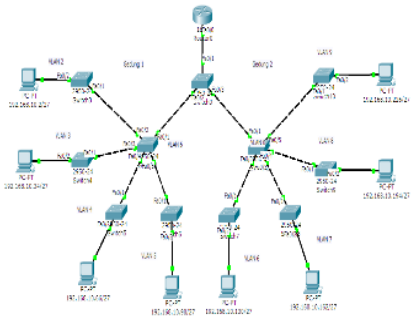
Few application of tree:

Linux: Linux is an backbone of the operating system like DEBIAN, UBUNTU, FEDORA and so on. And it supports many open source data science and big data applications.

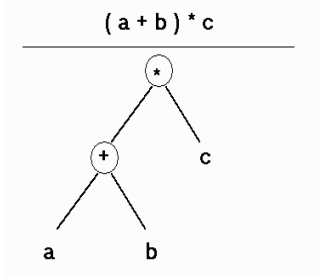
The most important thing is the Linux is also a tree structure.



NETWORKING ROUTE which also works on tree structure:



SYNTAX TREE in compiler: In compiler every expression, is converted into syntax tree



II.RECURRENCE RELATION

These are used to determine the running time of a programs.
Another aspect of recursion is its application to computer programming and maple programming.
In recursion comes the theory of automata which is concerned with the study of
mathematical models of computing devices.

III.PRIMITIVE RECURSION FUNCTION:

It can be constructed from the initial functions by a finite number of operations of
compositions and recursion only.

F(x,y+1)=x+(y+1)
=(x+y)+1
=f(x,y)+1=s(f<x,y>)

To show that f is computable, lets compute f(4,3) by using the recursion

F(4,0)=4
F(4,3)=s(f<4,2>)=s(s(f<4,1>))
=s(s(s(f<4,0>)))=s(s(s{4})))
=s(s({5}))=s({6})=7

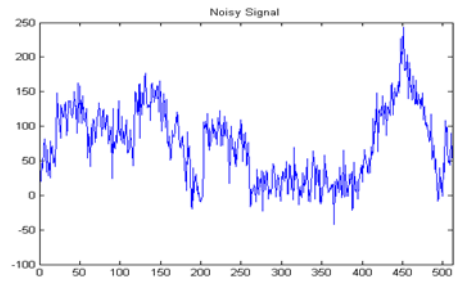
THEORETICAL COMPUTER SCIENCE which is a subset of general c.s and mathematics that focus more mathematical topics of computing and includes the theory of computation.

IV.TOPOLOGY

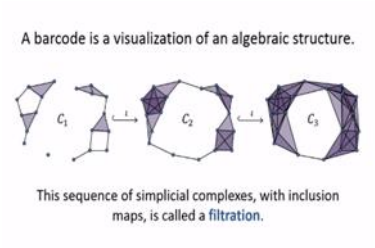
Topological data analytics(TDA):
In applied mathematics, **topological data analysis** (**TDA**) is an approach to the analysis of datasets using techniques from topology. Extraction of information from datasets that are high-dimensional, incomplete and noisy is generally challenging. TDA provides a general framework to analyse such data in a manner that is insensitive to the particular metric chosen and provides dimensionality reduction and robustness to noise.
The initial motivation is to study the shape of data. TDA has combined algebraic topology and other tools from pure mathematics to allow mathematically rigorous study of "shape".
The main tool is persistent homology, an adaptation of homology to point cloud data.
Persistent homology has been applied to many types of data across many fields.
Moreover, its mathematical foundation is also of theoretical importance.
The unique features of TDA make it a promising bridge between topology and geometry.

The popular tool in TDA is persistent Homology
PRESISTENT HOMOLOGY:

It is an algebraic method for discerning topological features of data
Topological data analysis and persistent homology have had impacts on Morse theory.
Morse work in persistent homology has extended results about Morse functions to tame functions or, even to continuous functions. A forgotten result of R. Deheuvels long before the invention of persistent homology extends Morse theory to all continuous functions. Topological data analysis and persistent homology have had impacts on Morse theory.

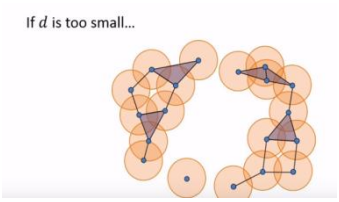


Morse theory has played a very important role in the theory of TDA, including on computation. Some work in persistent homology has extended results about Morse functions to tame functions or, even to continuous functions. A forgotten result of R. Deheuvels long before the invention of persistent homology extends Morse theory to all continuous functions.
Images. Can be represented by a function f : R 2 ! R
Data on manifolds. Can be represented by a function f : S 2 ! R

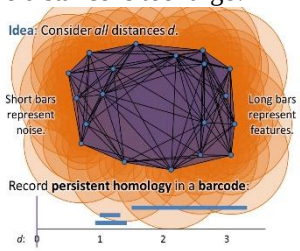


V.THE TOPOLOGICAL FEATURES THE FOLLOWING DATA EXHIBITS

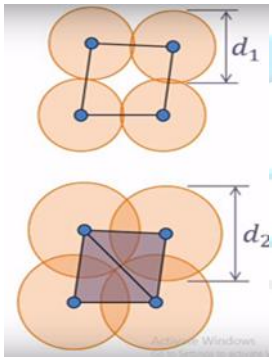
PROBLEM: how do we choose the distance?
If the distance d is too small then it gives hole which automatically creates noise:



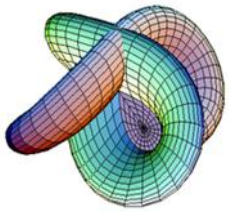
What if the distance is too large?



Each hole appears at a particular value of d a disappears at another value of d
We can represent the persistent of hole as a pair (d1,d2)
We can visualize this pair as a bar from d1 and d2. And the collection of such
Bars is called barcode.



VI.MATLAB



MATLAB (**mat**rix **lab**oratory) is a multi-paradigm numerical computing environment. A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python.

VII.MATLAB IN BIG DATA ANALYTICS:

Engineering and IT teams are using MATLAB to build today's advanced Big Data Analytics systems ranging from predictive maintenance and telematics to advanced driver assistance systems and sensor analytics. Teams select MATLAB because it offers essential capabilities not found in business intelligence systems or open source languages

VIII.WORKING WITH BIG DATA:

Tall arrays for big data:

Manipulate and analyse data that is too big to fit in memory

Create a tall array

```
tt = tall(ds)
```

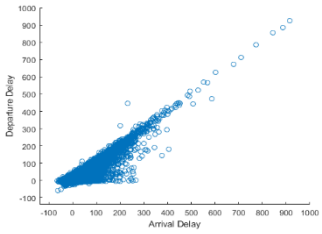
```
tt =
  M×5 tall table
      pickuptime          dropofftime          trip_distance    payment_type
    ____________________    ____________________    _____________    ____________
    2015-10-01 00:00:00    2015-10-01 00:05:48     1.1              2
    2015-10-01 00:00:00    2015-10-01 00:00:00    7.68              2
    2015-10-01 00:00:00    2015-10-01 00:00:00    2.53              1
    2015-10-01 00:00:00    2015-10-01 00:00:00     1.2              2
    2015-10-01 00:00:01    2015-10-01 00:16:19     3.8              1
    2015-10-01 00:00:01    2015-10-01 00:13:41     3.1              1
    2015-10-01 00:00:01    2015-10-01 00:21:23     4.5              2
    2015-10-01 00:00:01    2015-10-01 00:05:50    0.84              1
    :                      :                       :                :
    :                      :                       :                :
```

IX.BIG DATA ALGORITHMS

Perform support vector machine (SVM) and Naive Bayes Classification, creates bags of decision trees, and fit lasso Regression on out of memory data.

X.BIG DATA PLOT

Visualize out-of-memory data using plot, scatter, and binscatter.



XI.CONCLUSION

Mathematics has a solution to turn up the hard world of today and the advance study in mathematics is becoming a great problem solver in many fields and one such growth of mathematics is landed up in big data. This paper took out some idea of mathematics used in big data analytics and also tells that how mathematics had made easier the warm word of today –THE BIG DATA.

XII.REFERENCE

[1]. https://www.dmi.uns.ac.rs/mathbd/index.html
[2]. blog.press.princeton.edu/2012/04/.../how-mathematical-models-make-sense-of-big-da...
[3]. https://www.dmi.uns.ac.rs/mathbd/index.html