International Journal of Combined Research & Development (IJCRD)
eISSN: 2321-225X;pISSN:2321-2241 Volume: 4; Issue: 10; October -2015

Organizing Knowledge Extraction from Big Data Systems

Anirudh Prabhu,Rahul Shenoy
SIES-Graduate School of Technology,
Nerul, Mumbai University,India

***Abstract*—Even though some of the present-day technologies provide a number of solutions for handling large amounts of data, the increasing accumulation of data — also termed as Big Data — from the Internet such as emails, videos, images, and text as well as the digital data in medicine, genetics, and sensors and wireless devices is demanding efficient organizational and engi- neering designs. Many forms of digital data such as maps and cli- mate informatics, geospatial attributes such as global positioning coordinates, location information, and directions are represented by text, images, or interactive graphics-videos. A single source may produce various types of data (e.g. a geospatial data source may produce both image- and text-type data). This vast and rich data requires a generic processing mechanism that can adapt to various data types and classify them accordingly. In this paper, we propose a technique to optimize the information processing for on-the-fly clusterization of disorganized and unclassified data from vast number of sources. The technique is based on the fuzzy logic using fault-tolerant indexing with error-correction Golay coding. We present an information processing model and an optimized technique for clustering continuous and complex data streams. We show that this mechanism can efficiently retrieve the sensible information from the underlying data clusters. The main objective of this paper is to introduce a tool for this demanding Big Data processing — on-the-fly clustering of amorphous data items in data stream mode. Finally, we introduce the parallels between computational models of Big Data processing as well as the information processing of human brain where the human brain can be considered as a Big Data machine.**

I. Introduction

The new wave of technologies — such as social networks, geospatial computing, medical imaging, climate informatics, etc. — are enabling data to be generated, gathered and ana- lyzed at unprecedented scales. As a result, we can predict climate changes for future decades, sequence the entire human genome, identify the networking patterns in complex social networks, and leverage these results to solve complex real- world problems. But as the technologies improve in collecting accurate and minute data, processing and organizing that amorphous data with fault-tolerance is getting more and more complicated. To process 1 petabyte (PB) of data on a 1,000 node Amazon cloud, it would take 750 days and cost US

$6 million [1]. In this situation, heterogeneous and cloud computing will be unable to process the growing volume of digital data from vast number of resources. In spite of the tremendous advancements in computer technologies over the last few decades, when it comes to solving Big Data problems, however, the computational models for processing [3] remain the same as Von Neumann's computational model. Apart from introducing complexities, these models are quickly paralyzed when a large amount of data is communicated between computing systems [2].

The exponential growth in data and commercialization of parallel and cloud computing have given rise to some "Big Data" processing engines and large data processing appli- cations. But programmers face many challenges in parallel computing [3], and thus it is harder for the existing method- ologies and applications to process the large data sets faster and without introducing biases. Even if some methodologies manage to process large sets of data, the data sets for Big Data applications are often skewed. For example, [4] shows the data skew problem in the computation for cube materialization of large data sets and [5] faced the data skew problem in large-scale botnet detection. The database community should be focusing on addressing the persistent data skew problem in large data sets. Even though we can use some techniques from parallel database methodologies, many techniques are not efficient for the new large-scale setting.

One of the major fields that use Big Data models is medical sciences and it appears that the biomedical data problems have reached a stage when accumulation of more and more data does not produce outcomes commensurable with what is antic- ipated. The US Food and Drug Administration (FDA), medical research centers, and other organizations collect data on drug side effects from patients, drug companies, pharmacists, and physicians through FDA Adverse Event Reporting System [7]. This data provides unprecedented and valuable information about drug-related side effects. Unfortunately, those are often incomplete and biased, resulting in delays of life-saving alerts for the patients while data is being collected. One of the major challenges in these areas is that unclassified yet related data hide the adverse events resulting from multiple drugs interacting in unexpected ways [6].

Fig. 1 shows the unusual side effects by drugs on patients. Each patient cluster that is using two different drugs can have the same effects because of their drugs interacting in unexpected way. An analysis such as this can provide a deep insight into understanding the side effects and diseases. In large data sets, finding relationships between one cluster's attributes to another is not a simple task since the data is not organized or most of the time classified with significant biases.

Furthermore, mapping and constructing models of brain is another Big Data problem. The human brain is the most complex machine that we know. Currently Neuroscientists are trying to understand how the connections between neurons are made and how the data is transferred from one place to another

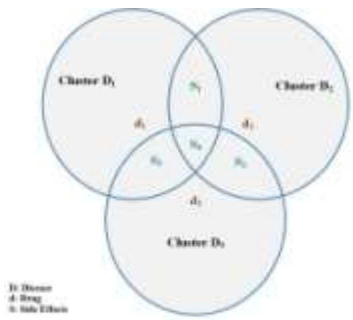

Fig. 1. Different Clusters of Patients with their Drugs and Side Effects

in the human brain. The human brain has 100 billion neurons connected by 100,000 billion synopses. Understanding how the brain works not only helps the medical field, but some of its data analysis and data retrieval techniques can be used in the real-world computer technologies [8]. It is not absolutely clear whether our computing models can imitate the computation model of the human brain since some of its functionality may be related to quantum world where the communication can happen instantaneously between two objects regardless of their distance — quantum entanglement. In any case, the organization of the brain can be considered as Big Data computational model.

Finally, when we look at the entire picture of how the data can be organized and classified, compared to direct analysis and sampling, clustering is the most efficient solution (see Fig. 2). Even in clustering, there are some drawbacks that have to be addressed [2]. When it comes to search, measuring the accuracy of the search algorithm is hard. For example, Google indexing technique — clusters of meta data (or meta knowledge) — can produce several false positive as well as false negative results due to obscure keywords. But access to ”Big Data” systems should be fault-tolerant keywords.

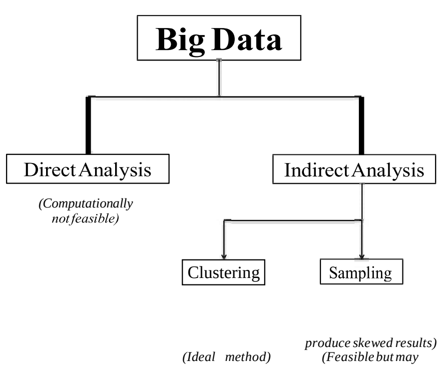

Fig. 2. Clustering is the Best Solution Available for Big Data Analysis

Our paper presents a new organizational procedure for clusterization of different types of data items such as texts, videos, images, etc. in data stream mode. This procedure uses perfect error-correcting Golay code model (23, 12, 7) to create 23-bit templates with 23 yes/no questions, which in turn creates a 12-bit data word with the hamming distance of 7. When each data item enters into the processing system, it is labeled with a 23-bit predefined questions template that can take 1-bit yes or no as answers (e.g. 1: Is it a text file? 2: ... 23: Does it contain an image?). Based on these answers, a unique binary label can be applied to the data item, which in turn efficiently allows the system to classify and cluster the data item accurately. Given the error-correcting properties of the code, it can correct errors up to 7 bits in the data word, which attributes to the high fault-tolerance for the procedure. This model could be an initial step towards achieving the process of analyzing very large collection of diverse information.

II. CLUSTERING WITH GOLAY CODING

Fuzzy logic based organization of information accessing is one of the complex problems in Computer Science. The solution of this problem is that a non-exact solution must be formulated by means of the exact computer interactions. Having many attributes for searching reduces the search complexity, initiates sequential lookups, and the desired speed up can be achieved by parallelization. In case if the faster access is the priority of the system then hashing technique can be used to fulfill the requirement. The most common approach to the fuzzy searching requires a hash transformation that can tolerate mismatches of bit attribute strings. This organization would allow the close by vector to be re-offered to the same hash table.

Assume that a medical data collection in a hospital has 23 different attributes such as diseases, drugs, side effects, etc. for each patient and there are a large number of patients. We further assume that the data of each patient can be compared or linked to other patients (e.g. connect to the same drug or disease). If the data is being collected from the patients constantly then each patient is characterized by a 23-bit codeword. For mapping the codeword we employ Golay code, which has the guarantee of having low hamming distance for their close decimal numbers in their binary representation. Our Golay code clustering algorithm is able to classify 23-bit codewords and organize them in a large number of clusters. The maximum Hamming distance from each of these clusters is 7 or 8 and for each cluster the total number of bit positions that have common bit values is either 16 or 15 [9]. This feature is especially important as the bit positions might have different physical properties, hence a low hamming distance alone does not mean that two codewords are same or similar. So the details of the patient may satisfy the features of different cluster and the group of patients with the same properties (e.g. diseases, prescribed medication, symptoms, side effects, etc.) can be easily identified with this clustering algorithm. Traditional clustering algorithms like k-means work on the Euclidian distance. Hence Golay code clustering have $O(n)$ time complexity and k-means have $O(n^2)$ time complexity.

In error-correction schemes, a number of parity bits are added to data to create a codeword. If some number of distortions happens during transmission, then the redundant parity bits can be used to restore the original data. The main concept of our clustering algorithm works by reversing the tradition error-correction schemes using the perfect Golay Code (23, 12, 7) (Fig. 3). ”Perfect” property constitutes the ability to cover all the possible outcomes. In case of Golay code, the total number of possible outcomes is 2^{23}. The entire set of 23-bit vectors — consider them as the vertices of 23-dimensional

International Journal of Combined Research & Development (IJCRD)
eISSN: 2321-225X;pISSN:2321-2241 Volume: 4; Issue: 10; October -2015

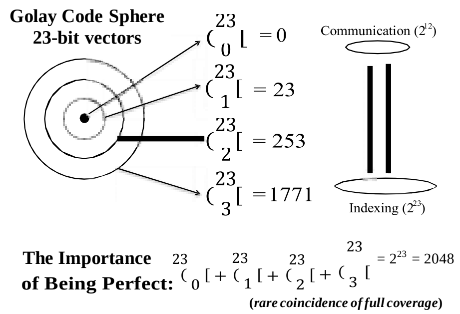

Fig. 3. Reversing the Error-correction Scheme with Perfect Encoding

binary cube — is partitioned into 2^{12} spheres. Each sphere has the radius of 3. Thus 23-bit strings (or bit streams) are mapped to 12-bit centers of these spheres giving the ability to tolerate some dissimilarities in the 23-bit stream. Thus our Golay code Big Data processing can exhibit high fault-tolerance that is vital for any large data streams to be processed on-the-fly.

III. FORMATION OF METADATA TEMPLATES

The implementation of this Golay code clustering has been described in [10]. In the design 26 letters of the English alphabet converted into 23 bit codewords showing the presence as well as the absence of a particular letter. In this implementation, the binary vectors preserve the closeness of the words with hamming distance of 2 in all of their operations from addition, substitution, and deletion of the word letters. We call it as FuzzyFind dictionary (Fig. 4). The developed dictionary for 100,000 words shows much better characteristics compared to the traditional clustering methods. This dictionary was very useful when it comes to categorizing vast catalog of digital documents to avoid improper indexing with wrongly spelled words.

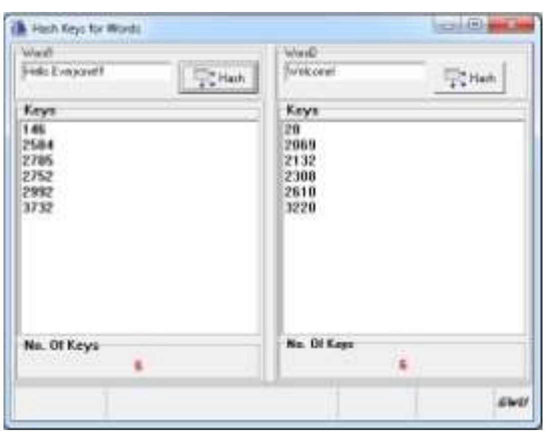

Fig. 4. Dictionary based on Fuzzy Search — Hash Key Generation for the Words

Hash is designed using the Golay code decoding procedure that takes neighborhood spheres of radius 1 around the 23-bit binary vectors and yields hash codes as 12-bit keys. 86.5% of the time it is mapped to to 6 hash indices and one hash index in 13.5% of the cases. So Golay code has a (large) sufficient data word space, which is 2^{12} and a large codeword space 2^{23}. Hence binary Golay code can be used for large data set clustering.

In some situations, 23-bit vectors at hamming distance 2 issue 12-bit indices, which have two common values. Hence, pairwise concatenation is used to place the data items with the binary vectors with the hamming distance 2 in the same cluster. One of the important advantages of using this scheme is to produce the required clustering effect [11]. This simple yet efficient pairwise indexing offers the effective collection and organization of unclassified and disorganized data. We consider a particular ontology approach called ”23-bit Metadata Templates”. This basically means that we introduce a template with 23 yes/no questions for each data item. The implementation model of this template is given in Fig. 5.

The scheme works in the following way: first, a group of ”23-bit Metadata Templates”, suitable for the given data set, is selected from a pre-established set of ”23-bit Metadata Templates”. Secondly, the collection of data items (or information items) will go through a sequential examination, and pointers to the data items are placed in an array with the pairwise addressing as the basis. So the method gives the result of storing the data items in corresponding clusters with each data item differing from another, no more than some limited number of bit-position mismatches of its 23-bit template.

Fig. 5. Extracting Knowledge from Amorphous Data

By using single bits, we can represent the presence or the absence of a certain attribute of the cluster corresponding to the data item. Several bits can be combined to represent numeric ranges or define the level of importance for each data item. Let us take an example: if clustering fails to identify an individual data item, other factors such as size of the file can play an important role in storing the data item in the correct cluster. Same as this, frequency of occurrence of an attribute or other features of a data item can be expressed in terms of some numbering techniques so that the clustering will not only be based on the presence or the absence of an attribute. These numeric quantities may be represented to characterize the attributes by selecting a precise coding technique to protect their closeness in the Hamming distance. The numeric representations may be used to represent some of the features of the attributes like the frequency of occurrence of a word in the document, number of pixels in a picture, or length of an audio file, etc.

International Journal of Combined Research & Development (IJCRD)
eISSN: 2321-225X;pISSN:2321-2241 Volume: 4; Issue: 10; October -2015

IV. DATA INTENSIVE COMPUTING WITH INFORMATION STREAMS

Error-correcting Golay code information processing, which necessary for the clustering of Big Data, requires continuous processing of intensive data streams on-the-fly. This kind of processing can be implemented through new pipeline architecture that is introduced in [12]. This structure provides a mechanism of effective solution for on-the-fly clustering from core to core. The pipeline has three basic components: Loading, Processing, and Unloading (Fig. 6). Loading is followed by processing that is followed by unloading:

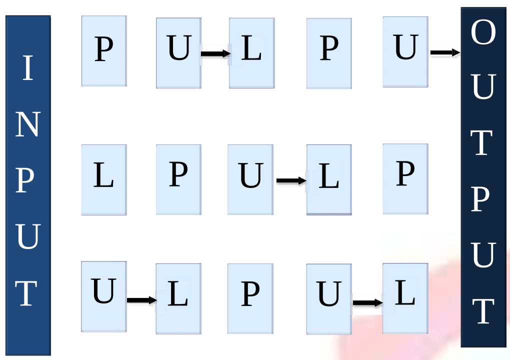

Fig. 6. On-the-fly Computations by a Microprocessor Pipeline (Loading, Processing, and Unloading Rotate in the Cycle: L→P→U→L)

This pipeline allows a procedure to be performed on- the-fly on a data item or a data set if there were enough processing cores. An important condition for the continuous data streams is realized by forced interrupts at each stage of the processing [12]. This structure uses data overlapping to address the problem of processed data sets across the classified segmentation created by the buffer size. With this feature, any algorithm that can work on a traditional microprocessor will be able to function in a continuous data stream setting. The data streams are partitioned and scheduled automatically to move through the pipeline.

This multi-processor architecture has another major advantage — it does not need the traditional complicated parallel processing methods. Instead this pipeline architecture can work well in the sequential code settings. Considering that the parallel programming in multiprocessing systems is becoming the biggest challenge for the industries, the proposed pipeline architecture uses dynamic resource management which in turn allows on-the-fly organization of data in intensive data streams. Hence the new architecture along with the Golay code clustering provides a strong framework for handling Big Data.

V. BIG DATA PROCESSING MODEL OF THE BRAIN

Our Big Data processing method can lead to an interesting direction where properly clustering massive amounts of data can empower and streamline decision-making process in computing systems. Here, the decision-making process would be operated under the framework of bounded rationality approach [13] — the more clusterized data you have in your system, the better the decisions of your system become. In a way, it is similar to the Google search algorithm where search results of a single query are affected by tens of thousands of data points such as users' history, search history, users' interest, current trends, etc. Google relies on billions of rows of metadata about the website indexed (or clustered) in appropriate servers. Based on this ideas, we can derive a computational model of biological Big Data machines such as the human brain. Understanding the information-processing model of the brain can lead to promising research in understanding and treating mental disorders (e.g. Autism and Depression).

The human brain consists of 10^{11} neurons and 10^{14} synapses. Formation of new connections and updates of the synapses are believed to be the base of the human memory formation. Human memory is a one-level system with non-erasable and virtually infinite memory capacity [18]. Yet our brain actively uses just 1 gigabit of information from the memory at any given time [19]. Despite the small memory information being active, our brain can process and retrieve information, and make decisions almost instantly — faster than any computer. Some of the studies show that babies wise up fast. By the time infants are 3 month-olds, their brain transforms from *tabula rasa* state to more mature by creating trillions of connections and the collective weight of all those neuron connections nearly triple in just a year. French neuroscientists say, when comparing adult and baby brain waves, the internal architecture seems to be in place to perceive objects in adult-like ways as early as 5 months [14]. Let us assume that these new connections and updates happen in $\frac{1}{100}$ of a second. At this rate to construct the neuron connection, it would take the system about 300,000 years with sequential processing. For a 3 year-old kind, it would take about 106 updates or new connections. We can compare this situation to printing vs. typing analogy where the algorithmic complexities of the processes are $O(1)$ vs. $O(N)$.

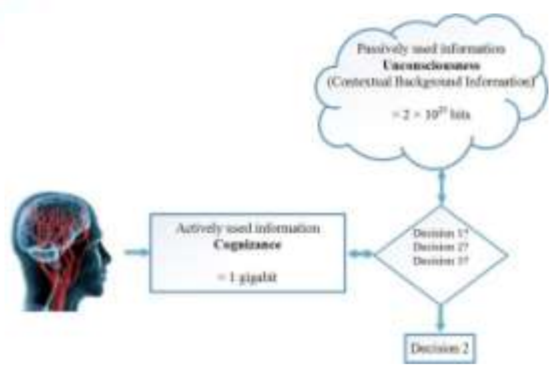

Fig. 7. Information Processing Model of the Brain in Big Data Framework

To understand the brain functionality, we need to examine it in terms of Big Data biology. Fig. 7 represents the information processing model of the brain. We can split the brain into two major parts in terms of Big Data biology, namely "Active" and "Passive". Active part participates in all processing operations i.e. cognizance; passive part presents a context that affects the selection of information for active part i.e., contextual background information or unconsciousness. This organization explains the ability of human brain to retrieve information and solve non-trivial problems.

This organization corresponds to Sigmund Freud's theory [17] — the effects of unconsciousness in human behavior — which is the founding theory of psychoanalysis. Freud's theory suggests that some unpleasant information contained in human unconsciousness can negatively affect the selections of "Active" part of the brain that governs the human behavior. Since human memory is non-erasable, Freud suggested that this situation may be cured by psychoanalysis — revealing the underlying unpleasant information and re-writing it in an acceptable context. This would affect the brain's "Active" part positively thus changing the human behavior.

The situation with autism deemed complimentary to Freud's scheme. The mind of developing child is missing some information in his unconsciousness that is responsible in a broad sense for various aspects of child's behavior. This condition produces inadequate behavioral response to certain circumstances. Therefore, to cure this disorder it is necessary to supply the child with appropriate "unconsciousness" information. This could be implemented by special algorithmic adaptations in a computer-game or virtual reality type setup introducing purposeful supervised selections from multiple proportions. The suggested curative actions will influence mostly the unconsciousness background.

A. *Bounded Rationality Approach to Artificial Intelligence*

In complexity theory, an Oracle Machine is an abstraction to study decisions problems. Oracle — a black box guiding the choice of available alternatives for a particular taska Turing computational model for Big Data processing. In other words, Oracle machine will be an expert machine to guide a through a particular process. Even though, P=NP — non-deterministic and deterministic problems are equal in their efficiency — is not solved, yet the computation with the Oracle machine has extraordinary influence and efficiency. The Big Data computational framework exhibits the properties and other features of an Oracle machine.

When we consider the brain's information processing in Big Data framework, a major problem of artificial intelligence — making right decisions with the huge dataset — rise in either real or synthetic data. The Oracle machine tackles this problem with context-addressable access. Context-addressable means the context of a data set to which the information is supplied.

Using context-addressable access — a simplified model of the system — , which is guided by an "Oracle Machine", can be designed. An exemplar of an "Oracle" could be produced within the suggested computational model using a rich context of Big Data. So, this "Artificial Intelligence" system could acquire "Intuition from Context" (Fig. 8). One of the classical Artificial Intelligence applications is the game of chess. Here an Oracle machine can be a qualified human expert or a supercomputer. The beginner would learn from his/her mistakes as well as feedback from the expert or supercomputer thus a substantial improvement can be achieved. In Big Data term, it as writing new information in the beginner's memory. This approach can be used in autism treatments.

Just like Big Data information processing (Fig. 8), several types of algorithmic systems can be designed specific to the

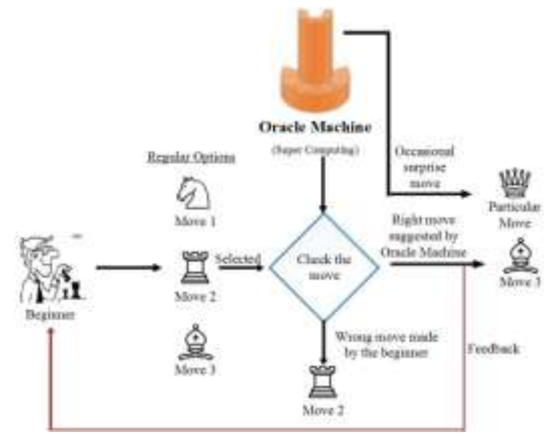

Fig. 8. Big Data Information Processing — a Chess Game

type of autism spectrum disorder (ASD). In this case, the supercomputer will be an expert psychiatrist or predefined supervisor program to guide the autistic children who are undergoing this procedure. We will be able to design several scenarios based on the type of ASD (lagging in social interactions, communication, behavior, and interests) for the autistic children.

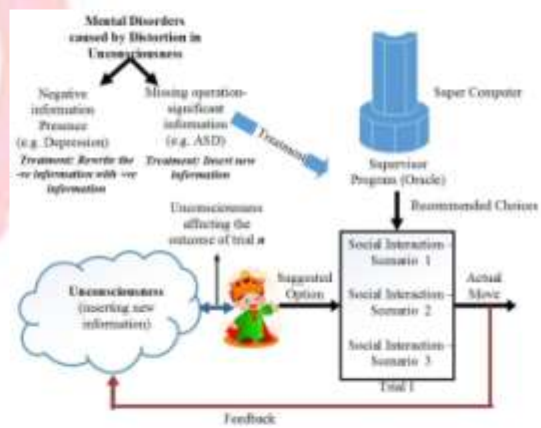

Fig. 9. Treating Autism Spectrum Disorders

With the suggested approach we are planning to test another scheme: a beginner displays several possible scenarios in accordance with some simplified understanding of the program and an "Oracle Machine" (a qualified psychiatrist or a supervisor program) makes a best selection of the displayed moves (Fig. 9). The oracle machine provides feedback and presumably it will be affecting the unconsciousness (contextual background information). After a certain period of time, we can start to see the improvements in autistic child's behavior. Some of the studies are confirming by rare cases of spontaneous recovery. Fig. 9 is an illustration of a treatment trial for social interaction type ASD. This treatment mechanism is highly flexible, adaptable and patient-centered algorithmic approach. By having a supervisor program, the treatment becomes more affordable thus allowing children from poor families to get the treatment.

B. Effects of Contextual Background Information or Unconsciousness

The influence of unconsciousness in the human behavior is far greater than previously thought. Psychological scientists still perceive the unconsciousness (or unconscious mind) as the ”real” shadow of the conscious brain. The effects of unconscious mind are perceived differently from cognitive psychology to social psychology. Collective results from the cognitive psychology suggest that the role of unconscious mind is limited. Yet the results from social psychology suggest the unconscious mind is a pervasive and powerful influence over such higher mental processes [15]. In evolutionary biology perspective, the unconsciousness plays a vital role in adaptive self-regulation of human brain. This means that the decisions made and the non-trivial problems solved by the human brain are indeed influenced by this contextual background information.

One of the fundamental models for unconsciousness was provided by Sigmund Freud’s theory of effects of unconsciousness in human behavior. This theory still influences many scientists research, especially outside of psychological science. Freud’s model of the unconscious brain is fundamental and detailed than any other model found in cognitive or social psychology. But experimental evidence collected through empirical tests was not kind to the specifics of Freud’s theory. Although the evidences from cognitive psychology and social psychology does not support Freud’s theory in general terms, its historic importance is great and it’s still the best model for the unconsciousness.

The theories based on unconsciousness suggest that past events stored away in human brain can actively impact their current actions even if the individuals are not aware of this effect. Depression, for example, caused by many factors including abuse, death or a loss, major events, conflicts, serious illnesses and other personal problems. Since human memory is non-erasable these major events or illnesses are deeply stored in the brain i.e. in unconscious. These memories in the background influence all major thought processes in our active brain — negatively affecting the human behavior. Even though people with manic-depression know their problem, they have very little control over this excessive negative contextual background information.

When it comes to autism, more than 500 genes so far have been implicated in causes of autism showing no clear relation to the genetic causes [16]. So we consider autism in the Big Data framework — a distortion in the growing child’s brain. More or less, some types of autism are the opposite of depression scenario: excessive unwanted information in contextual background causes depression and low or no presence of necessary information in the contextual background causes autism. The mind of developing child is missing some information in his unconsciousness that is responsible in a broad sense for various aspects of child’s behavior. This condition produces inadequate behavioral response to certain circumstances.

VI. Conclusion

In this paper, we proposed a generic Big Data information processing model based on perfect error-correcting codes — Golay Codes. We presented a fuzzy find dictionary that proves the hash mapping to be efficient in searching as well as storing the Big Data streams on-the-fly. We also presented a Big Data information processing model based on bounded rationality approach (extrapolated from Big Data processing model). Based on this Big Data model, we were able to derive the information processing model of biological Big Data machines such as the human brain, which lead to the development of efficient treatment mechanisms for autism and other mental disorders. Most of all, we presented a Big Data processing approach that can be used for massive data collection and processing in geospatial research. Our Big Data processing model outperforms the existing clustering methods with $O(N)$ time complexity. In the future, we will extrapolate our theory to the field of medicine where we can analyze the impact of various drugs and their combinations. This may lead to the finding of interesting patterns and dynamics of human health.

Acknowledgment

We would like to thank the chairman of Computer Science Department of George Washington University Prof. Abdou Youssef for helping us by providing office space and equipment for the research. We also would like to thank Prof. Zhiping Pang of University of Medicine and Dentistry of New Jersey (UMDNJ) for his inputs on developing the brain information-processing model.

References

[1] O. Trelles, P. Prins, M. Snir, and RC. Jansen. *Big data, but are we ready?*, Nature Reviews Genetics 12.3 (2011): 224-224.

[2] S. Berkovich and D. Liao,. *On clusterization of big data streams.*, Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications. ACM, 2012

[3] G. Mani, S. Berkovich, and I. Mihai. *A combinatorial distributed architecture for exascale computing*., In 2012 Fourth International Conference on Advanced Computing (ICoAC),(pp. 1-5). IEEE.

[4] E. Gonina, A. Kannan, J. Shafer, and M. Budiu. *Parallelizing large-scale data processing applications with data skew: a case study in product-offer matching*., Proceedings of the second international workshop on MapReduce and its applications. ACM, 2011.

[5] Y. Zhao, Y. Xie, F. Yu, Q. Ke, Y. Yu, Y. Chen, and E. Gillum. *Botgraph: large scale spamming botnet detection*. In Proceedings of the 6th USENIX symposium on Networked systems design and implementation, pages 321-334, Berkeley, CA, USA, 2009. USENIX Association.

[6] R. White, N. P. Tatonetti, N. H. Shah, R. B. Altman, and E. Horvitz. *Web-scale pharmacovigilance: listening to signals from the crowd*, in J Am Med Inform Assoc. 2013 Mar 6.

[7] *FDA Adverse Event Reporting System*, http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation /Surveillance/AdverseDrugEffects/default.htm

[8] The Human Brain Project: A report by European Commission, http://www.humanbrainproject.eu/files/HBP_flagship.pdf

[9] H. Yu, S. Berkovich, Ta. Jing, and D. Chen. *Golay Code Clustering for Mobility Behavior Similarity Classification in Pocket Switched Networks*, In WASA 2011, pp. 302-310.

[10] S. Berkovich, E. Berkovich, B. Beroukhim, and G.M. Lapir, *Organization of Automatic Spelling Correction: Towards The Design of Intelligent Information Retrieval Systems*, The 21st National Conference of the ASEM, Washington, DC, Oct. 2000, pp. 525-527.

[11] H. Yu, T. Jing, D. Chen, and S.Y. Berkovich, *Golay Code Clustering for Mobility Behavior Similarity Classification in Pocket Switched Networks*, J. of Communication and Computer, USA, No 4, April, 2012.

[12] D. Liao and S.Y. Berkovich, *A New Multi-Core Pipelined Architecture for Executing Sequential Programs for Parallel Geospatial Computing*, in Proceedings of the 1st international Conference on Computing for Geospatial Research & Application, Washington, D.C., June 21 - 23, 2010. COM.Geo '10, ACM, New York, NY, U.S.A., 2010.

[13] L. Cosmides and J. Tooby, *Better than rational: Evolutionary psychology and the invisible hand*, American Economic Review 84, no. 2 (1994), 327-332.

[14] G. Mohan, *Adult consciousness may be present in infants' minds, study says*, Los Angeles Times, April 19, 2013.

[15] J.A. Bargh, and E. Morsella. *The unconscious mind*., Perspectives on psychological science 3, no. 1 (2008): 73-79.

[16] R. Bazell, *Experts: Wide 'autism spectrum' may explain diagnosis surge*, http://vitals.nbcnews.com/_news/2012/03/29/10925941-experts-wide-autism-spectrum-may-explain-diagnosis-surge?lite

[17] S. Freud. (1922). *The unconscious*. The Journal of Nervous and Mental Disease, 56(3), 291-294.

[18] R. L. Greene. (2014). *Human Memory: Paradigms and Paradoxes*. Psychology Press, 2014, 1-51.

[19] S. Berkovich, *Formation of Artificial Intelligence and Natural Intelligence in Big Data Environment*, Network Science and Cybersecurity 55, 2013, 189.