

A Theoretical Framework for Big Data Analytics Based on Computational Intelligent Algorithms with the Potential to Reduce Energy Consumption



Haruna Chiroma, Usman Ali Abdullahi, Ibrahim Abaker Targio Hashem, Younes Saadi, Rawaa Dawoud Al-Dabbagh, Muhammad Murtala Ahmad, Gbenga Emmanuel Dada, Sani Danjuma, Jaafar Zubairu Maitama, Adamu Abubakar and Shafi'i Muhammad Abdulhamid

Abstract Within the framework of big data, energy issues are highly significant. Despite the significance of energy, theoretical studies focusing primarily on the issue of energy within big data analytics in relation to computational intelligent

H. Chiroma (✉) · U. A. Abdullahi
Department of Computer Science, Federal College of Education (Technical), Gombe, Nigeria
e-mail: freedomchi@yahoo.com

U. A. Abdullahi
e-mail: danzazzau12@gmail.com

U. A. Abdullahi
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Perak, Malaysia

I. A. T. Hashem
Centre for Data Science and Analytics, School of Computing & Information Technology, Taylor's University, Subang, Jaya, Malaysia, Kuala Lumpur, Malaysia
e-mail: ibrahimabaker.targiohashem@taylors.edu.my

Y. Saadi
Department of Computer Science, University of Batna, Fésdis, Algérie
e-mail: younessaadi@gmail.com

R. D. Al-Dabbagh
Department of Computer Science, University of Baghdad, Baghdad, Iraq
e-mail: rawaaiq04@yahoo.com

M. M. Ahmad
Department of Information Technology, National Open University of Nigeria, Lagos, Nigeria
e-mail: mursuw@gmail.com

G. E. Dada
Department of Computer Engineering, University of Maiduguri, Maiduguri, Nigeria
e-mail: gbengadada2004@yahoo.com

S. Danjuma
Department of Mathematical Science, North-West University Kano, Kano, Nigeria
e-mail: sani_danjuma@yahoo.com

© Springer Nature Switzerland AG 2019
T. Herawan et al. (eds.), *Advances on Computational Intelligence in Energy, Green Energy and Technology*, https://doi.org/10.1007/978-3-319-69889-2_1

algorithms are scarce. The purpose of this study is to explore the theoretical aspects of energy issues in big data analytics in relation to computational intelligent algorithms since this is critical in exploring the empirical aspects of big data. In this chapter, we present a theoretical study of energy issues related to applications of computational intelligent algorithms in big data analytics. This work highlights that big data analytics using computational intelligent algorithms generates a very high amount of energy, especially during the training phase. The transmission of big data between service providers, users and data centres emits carbon dioxide as a result of high power consumption. This chapter proposes a theoretical framework for big data analytics using computational intelligent algorithms that has the potential to reduce energy consumption and enhance performance. We suggest that researchers should focus more attention on the issue of energy within big data analytics in relation to computational intelligent algorithms, before this becomes a widespread and urgent problem.

Keywords Big data analytics · Energy · Cluster systems · Computational intelligent algorithms · Artificial neural networks · Cuckoo search algorithm

1 Introduction

The International Energy Agency (IEA) has estimated that the global consumption of energy is expected to surge by 53% by the year 2035 [1]. Energy is viewed as the largest industry across the globe [2]. The consumption of energy involves all sectors of society, including information and communication technology. Shojafar et al. [3] have argued that real-time processing and energy efficiency are hot topics in the management of information and communication technology platforms. Currently, one of the most widely discussed topics in the science and technology community is big data. Big data has potential for applications in all sectors of society, such as climate, economics, health, social science [4]. The data collected from various sources in society is growing exponentially and is estimated to grow to 44 ZB (trillion gigabytes) by 2020, from 4.4 ZB in 2013 [5]. In big data, a diverse field of study which includes natural language processing, medical science, secu-

J. Z. Maitama

Department of Information Technology, Bayero University Kano, Kano, Nigeria
e-mail: jzmaitama.it@buk.edu.ng

A. Abubakar

Department of Information Systems, International Islamic University Malaysia, Gombak, Malaysia
e-mail: 100adamu@gmail.com

S. M. Abdulhamid

Department of Cyber Security Science, Federal University of Technology Minna, Minna, Nigeria
e-mail: shafii.abdulhamid@futminna.edu.ng

riety and business management depends heavily on knowledge discovery through big data analytics. The effective and efficient processing of big data requires computer systems [6] involving Hadoop which offer the MapReduce framework for parallel computation [7].

The transfer of large volumes of data between users, service providers and data centres requires a high bandwidth connection. This consumes large amounts of energy, more than simply processing and storing the big data within cloud-based data centres. Therefore, emits high carbon dioxide. The transfer of big datasets into remote data centres consumes a significant quantity of power [8], and these carbon dioxide emissions contribute to global warming [9]. The optimisation of energy consumption for data transmission requires the network to reduce redundant and duplicate traffic [10].

Applications for future generations of parallel and distributed systems in big data analytics are a major issue. These applications generate datasets in repositories that exceed exabytes, and the size of these datasets is speedily increasing. These datasets and their associated applications pose a challenge to both software techniques and software development [11]. The task of analysis frequently has strict targets, and one of the major issues for applications in this field is the quality of data. Most of the emerging applications, data-driven models and techniques with the capability of operating at large scales are not yet widely known [12]. In real-time systems, the amount of energy is increasing; thus, the application of big data methodologies can be used to handle these operations [13]. Significant developments in big data have arisen from various research communities, for example data mining and learning algorithms from the artificial intelligence research community [4].

Big data offers important opportunities for organisations who can analyse it and gain critical intelligence for effective decisions [14]. Within various industries, data processing and analysis plays a significant role, especially in situations where non-linear dynamics, comprising various uncertainties and mathematical models, fails. Computational intelligent algorithms such as fuzzy logic, artificial neural networks and evolutionary algorithms have demonstrated their ability to deal effectively with data modelling, and research on computational intelligent algorithms has attracted unprecedented attention from researchers. Computational intelligent algorithms have successfully solved real-world problems, as reported in the literature. Examples of the effectiveness of computational intelligent algorithms in solving real-world problems include control engineering, the modelling of unknown nonlinear dynamics using artificial neural networks and the implementation of controllers using adaptive neuro-fuzzy inference systems. Despite the overwhelming successes recorded by computational intelligent algorithms in solving real-world problems, those within the research community are still facing obstacles to the processing of industrial data, such as feature extraction from large-scale measurements that are distributed in nature, machine learning algorithms for highly robust data modelling and signal processing [15].

The world is experiencing a data revolution in terms of gleaning knowledge from big data. Computational intelligent algorithms are among the mainstream tools of big data analytics: computational intelligence has played an important role in artifi-

cial intelligence, which focuses on the design of algorithms. Such algorithms can be used to analyse huge amounts of structured and unstructured data which help in the discovery of approximate solutions for many complex problems [16]. Meanwhile, the use of virtualised clouds is currently under consideration in big data analytics, based on new machine learning theories and artificial intelligence. It is now common that intuitive physical interpretations affect the use of machine learning and artificial intelligence. It is therefore important that a suitable knowledge interpretation is provided, in order to make sound decisions based on the intelligence derived from machine learning and artificial intelligence [17].

In this chapter, we propose a theoretical framework for big data analytics, based on computational intelligent algorithms with the potential to reduce energy consumption and improve performance. It is necessary to explore the theoretical aspects of energy issues in big data analytics in relation to computational intelligent algorithms; it is critical to explore the theoretical aspects of big data in view of the fact that this can point the way towards effective and efficient applications [4].

The remaining sections of this chapter are organised as follows: Section 2 introduces the computational intelligent algorithms; Sect. 3 presents a discussion of big data analytics and the energy consumption of cluster systems; Sect. 4 discusses big data analytics and computational intelligent algorithms; Sect. 5 describes the issue of energy consumption in the application of computational intelligent algorithms in big data analytics; Sect. 6 describes the proposed framework for big data analytics based on computational intelligent algorithms; and Sect. 7 presents the conclusions.

2 Computational Intelligent Algorithms

The computational intelligent algorithm is a name recently given to the branch of artificial intelligence that deals with sub-symbolic techniques. It offers a description of techniques that mainly focus on strategies and results. Computational intelligent algorithms include sub-disciplines that deal with adaptive and intelligent systems such as evolutionary computing, artificial neural networks, fuzzy systems, artificial immune systems, particle swarm optimisation (PSO) and ant colony optimisation [18, 19]. The primary source of inspiration for these intelligent systems is nature; most of these algorithms are inspired by the characteristics of biological systems, referred to as biology-inspired algorithms [20].

Solving real-world problems generally involves challenging and NP-hard problems which require optimisation techniques, with no guarantee of obtaining an optimal solution. There are no effective and efficient algorithms for all NP-hard problems; therefore, experimentation with various optimisation algorithms is required to find the algorithm that produces the optimal solution. Many computational intelligent algorithms such as PSO, cuckoo search and firefly have been introduced to deal with the challenges of optimisation problems [20]. Computational intelligent algorithms have become widespread, and this has significantly expanded the literature [21].

Recently, a new computational intelligent algorithm inspired by nature has been added to the literature almost every month. It is likely that there are more than 200 of such algorithms in the literature. As these algorithms have flooded the literature, many researchers have found that the newly created algorithms are the existing algorithms disguised as new ones [22]. Figure 1 illustrates the number of computational intelligent algorithms introduced into the literature per year. In 2009, the literature witnessed a drastic influx of computational intelligent algorithms. Figure 2 presents the classification of computational intelligent algorithms inspired by nature and is based on the classification proposed by Fister et al. [20].

2.1 Characteristics of Computational Intelligent Algorithms

In general, computational intelligent algorithms aim to generate a new solution which is superior to the existing one. Ideally, computational intelligent algorithms are expected to generate solutions superior to current solutions with minimal effort [21]. We now examine the major characteristics of computational intelligent algorithms: exploitation and exploration, parameter tuning/control, diversity and adaptation and algorithm parameters.

Exploitation and Exploration Exploitation is a local search process using local information for a problem and uses information obtained from a problem to generate new

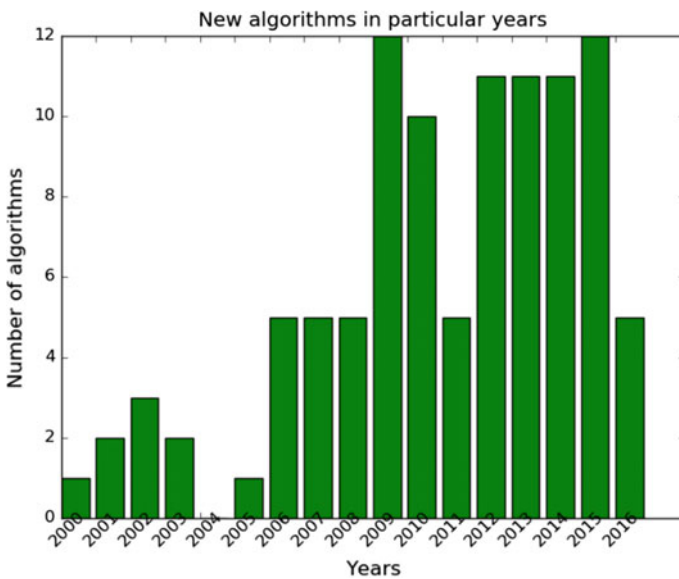


Fig. 1 Number of computational intelligent algorithms in the literature per year

solutions which are superior to the existing ones. The major strength of exploitation is its ability to give a high convergence rate. However, it has the possibility of becoming stuck in local minima. Conversely, exploration is a global search process which allows computational intelligent algorithms to explore the larger search space in far regions efficiently; it has the ability to generate solutions with sufficient diversity which is far from the existing solutions. Exploration has a lower propensity to become stuck in local minima, but it has a slow convergence rate and involves a high computational cost. Good performance for an algorithm requires a balance between exploitation and exploration: high exploitation and low exploration lead to faster convergence, but the possibility of finding a true global solution is low, while low exploitation and high exploration can lead to the meandering of the search path with a slow convergence rate [21].

Parameter Control/Tuning The values of parameters obtained through parameter tuning are fixed during iterations, whereas the parameters of an algorithm are varied for the purpose of control. Parameter control aims to find the algorithm with the best convergence rate for better performance; parameter tuning is carried out to find the optimal parameter settings for the running of the algorithm, in order to solve a broader array of problems. There is currently no systematic and efficient method of tuning to obtain optimal parameter settings; this is often realised through extensive experiments on parameter studies [23].

Diversity and Adaptation The computational intelligent algorithms have both diversity and adaptation, which are evident from the balance between exploitation and exploration. For example, ways of balancing exploration and exploitation are the key form of adaptation. For instance, the representations of solutions in genetic algorithms are either in binary or in real number form, whereas swarm intelligence-

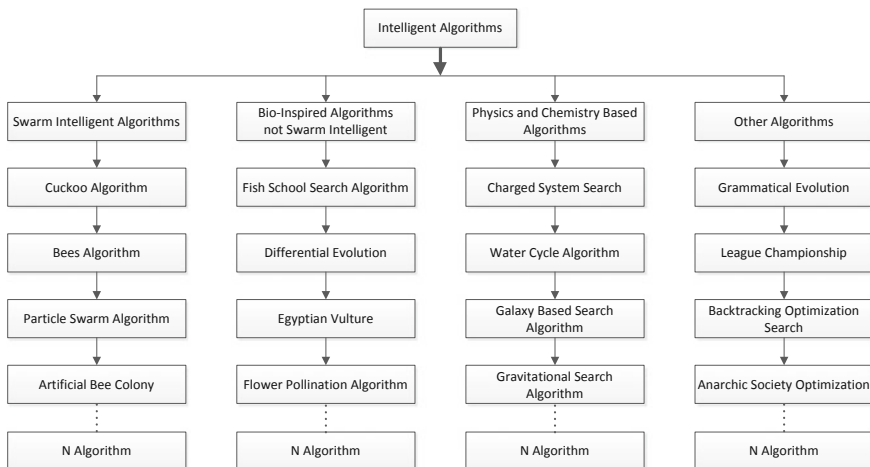


Fig. 2 Classification of intelligent algorithms

based algorithms generally use real numbers for solutions. The population size of an algorithm can either be fixed or varying, and variation in population size is therefore a typical example of adaptation.

Algorithm Parameters These algorithms involve parameters, and algorithm operators are used to construct the algorithms. In genetic algorithms, crossover, mutation and selection are used. Crossover is the operation used to create new solutions [21]. As an example of the differences between computational intelligent algorithms in terms of parameters, strengths, weaknesses, generation of new solution and solution representation, Table 1 presents five different well-established algorithms from the literature with their differences.

3 Big Data Analytics and Energy Consumption by Cluster Computing Systems

3.1 Big Data Analytics Platforms

The unprecedented accumulation of data in the information technology world has given rise to the concept of big data. Volumes are extremely large, with petabytes (PB) and even zettabytes (ZB) of data handled by organisations. The velocity and time-based variability of this data involve high speeds. The formats in which this data is created and stored are inconsistent, although these may originate from the same source and/or be generated by the same user. The veracity of the data, as opposed to the noise inherent in it, is of the highest concern. Despite all these features, big data offers high value when properly stored and analysed [29, 30]. Internet companies handle large volumes of Internet requests from their users using big data analytical platforms running on clusters of commodity hardware. Facebook and Walmart are two good examples [31].

Due to the above-mentioned characteristics, the storage and analytics of big data require large hardware resources. For example, in order to store 1 PB of data on a cluster with a 6 TB capacity hard disk, 163 hard-disk units are required. Assuming each node in the cluster can host five hard-disk units, then a 100-node cluster is required. In addition, when analysing this 1 PB of data, each hard disk must be accessed for IO, depending on the platform used for big data analysis. This is due to the fact that big data analytics platforms use the sequential access method by default. The most popular big data analysis framework, Hadoop, uses a full scan (sequential access) by default for the targeted data [32]. Each IO operation on a hard disk leads to a maximum consumption of energy in the form of electricity. *Consumption due to powering of the cluster and cooling:* The scenario described above can be used to give a wider picture of how much energy is consumed by companies that deal with big data. This explains why power and energy are always of first-order priority in the design of computing systems infrastructure [31].

Table 1 Major differences between five well-established computational intelligent algorithms

Algorithm	Strengths	Weaknesses	Generation of new solutions	Parameters required for settings	Solution representation	Reference
CS	Strikes a balance between exploitation and exploration. Requires few parameter settings	Sensitive to parameter settings	Lévy walk and far-field randomisation	Probability of the worst nest and step size	Real number	Yang and Deb [24]
FP	Balance between consistency and exploration	Tuning switching probability can create fluctuations	Pollination	Switching probability	Real number	Yang [25]
ABC	Strong exploration	Weak exploitation. Requires many parameter settings	Randomness	Number of scouts, swarm size, number of onlookers and number of employed bees	Real number	Karaboga [26]
PSO	Computationally inexpensive and very simple to implement	Possibility of premature convergence	Adding velocity to a position	Population size, probability, gamma and neighbourhood	Real number	Eberhart and Kennedy [27]
GA	Strong exploitation	Unnecessary number of iterations. Weak exploitation	Crossover	Probability of crossover, probability of mutation, population size, generations	Real number/binary number	Holland [28]

CS Cuckoo search algorithm, *GA* genetic algorithm, *FL* flower pollination algorithm, *ABC* artificial bee colony

Initially, information technology companies used clusters of commodity hardware and networking to avoid the high costs of hardware. These companies see this as a better option for providing the infrastructure necessary to accumulate and process large amounts of data, in comparison with the expense of purchasing and maintaining supercomputers and mainframe computers which would allow the system to achieve the same purpose. However, the amount of power and energy consumed by these clusters, particularly during big data storage and analysis, is growing at an alarming rate. The cost of the energy used by servers within their lifetime is expected to supersede the cost of the hardware itself, if current trends continue unchecked [31]. For example, Yahoo has installed a Hadoop cluster of over 2000 servers, while that of Facebook has more than 600 servers. Similarly, General Electric has deployed Hadoop on a cluster of 1700 servers. Energy is a crucial issue in view of these massive deployments of Hadoop over thousands of cluster systems and has influenced the cost of exploiting cluster systems. In 2007, there was a high cost of energy consumption for cluster systems [33]. This example shows the massive deployment of Hadoop over a cluster of hundreds of servers. In addition, the deployment and operation of Hadoop, the hardware required to build the cluster systems and the energy required to run them incur very high costs [31].

3.2 *Energy Consumption Over Big Data Platforms*

Energy consumption in big data platforms is related to several factors such as physical resources and computing resources. Big data platforms have the ability to model, organise, store and process large amounts of data. Development of information technology platforms and the massive generation of data in the world, big data technologies has become the battlefield of information technology service providers in terms of high performance and cost. The research community started to focus on energy consumption in big data platforms [34]. With the fast development of the global economy, energy consumption will keep increasing in the upcoming years.

Figures 3 and 4 are created based on the data presented in [35]; it shows the carbon dioxide emissions of data centres according to [36]. Whereas Fig. 4 depicts the distribution of energy consumption in data centres, Figure 3 shows that almost 6% per year of the emissions are caused by information technology servers. The consumption of energy in big data platforms can be expressed as follows:

$$E_c = C_r + P_r \quad (1)$$

where E_c denotes the energy consumption, C_r represents the computing resources and P_r denotes the physical resources. According to [37], the energy consumption of computing resources accounts for about 50% of the total energy consumption as shown in Fig. 4. The percentage of energy consumes by servers' computation; the communication equipment, and the storage devices are depicted in Fig. 4.

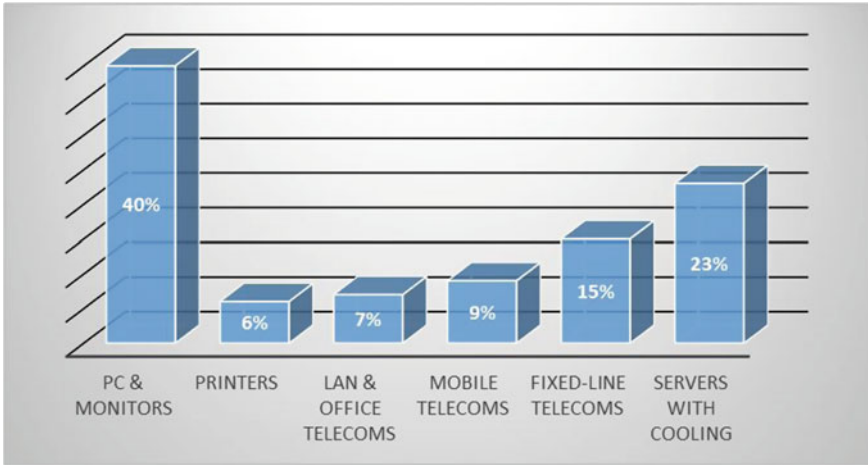


Fig. 3 Estimated information and communication technology carbon dioxide emissions

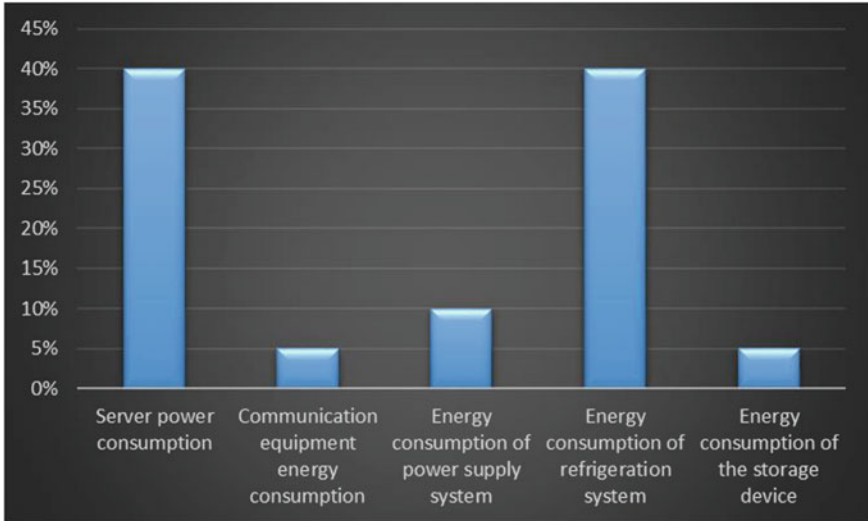


Fig. 4 Energy consumption distribution of data centres

From Fig. 4, it is clear that the data servers are consuming a big part of the energy consumed by data centres. This amount grows exponentially in case of processing large datasets which is the case of big data platforms. Therefore, reducing energy consumption for big data platforms is the key issue for sustainable big data platforms.

3.3 Metrics Used for Measuring Power in Big Data Platforms

The management of energy consumption can be formulated as a multi-objective optimisation problem, where several performances and energy metrics are used to measure the performance [38–42]. It is very important to highlight that the following two objectives are mostly considered in the literature: minimisation and maximisation. The minimisation consists of reducing the consumption of the data platforms during peak power. The maximisation consists of increasing energy efficiency.

In fact, limiting the consumption during peak power is very crucial to maintain the reliability of big data platforms, escape system overheating and avoid power capacity overloads. It is shown that reducing power consumption is strongly correlated with the cost of power provisioning [43]. Energy efficiency can be expressed as follows:

$$\text{Energy efficiency} = \frac{\text{Computing performance}}{\text{Total energy consumed}} \quad (2)$$

This metric represents the main focus of energy management of data centres and processing systems. From the perspective of power management of data centres, energy consumption control is viewed as a result of [39]:

Power Usage Effectiveness Facility efficiency is the ratio of the total amount of energy used by a data centre facility to the energy delivered to computing equipment.

Server Power Usage Effectiveness Server power conversion efficiency is the ratio of the total server input power to its useful power consumed by the electronic components directly involved in the computation.

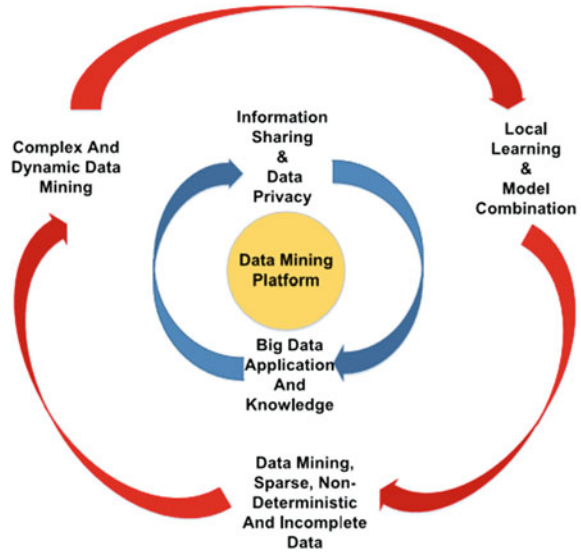
Server's Architectural Efficiency Server's architectural efficiency is the ratio of computing performance metric to the total amount of energy used by electronic components.

4 Computational Intelligent Algorithms and Big Data Analytics

In today's world, almost everything is online, and organisations intending to improve their services analyse big data to gain knowledge to be used in improving their services [44]. Big datasets are beyond the scope of relational or object-oriented databases, and traditional computer applications and normal computers cannot handle the analytics involved. These big datasets require very large parallel processing power, from clusters of computers, for analysis. The processing of big data is generally based on nonlinear systems, and actions are not predictable in some cases [45].

To discover the knowledge required for decision making, data mining algorithms are applied to the datasets extracted from data sources. In recent years, much attention

Fig. 5 Three primary layers of big data analytics [49]



has been given to data mining, probably due to the popularity of big data concepts [44]. Big data analytics involves modelling, analysis and interpretation [46]. It has been shown that computational intelligent algorithms can be applied to solve big data problems effectively from the perspective of hardware and software design [29].

The application of computational intelligent algorithms in big data analytics is severely limited, however, since recent intelligent algorithms have difficulty in analysing big data. This is because the nature of big data makes it difficult for these intelligent algorithms to analyse it [47]. The proposed basic framework for big data analytics in relation to data mining is shown in Fig. 5; these data mining algorithms also include computational intelligent algorithms [48] (discussed in Sect. 2).

The commonly accepted framework for big data analytics is shown in Fig. 5. It comprises three layers [50] as follows:

- i. Data access and computing,
- ii. Data privacy and
- iii. Domain knowledge and data mining algorithms.

The core of Fig. 5 is the data mining platform, which is responsible for data access and computing processes. With the increasing accumulation of high volumes of data, the distributed storage of large-scale data is required to be considered during computation. In brief, data analytics and the processing of the task are partitioned into sub-tasks in multiple forms for parallel execution on a large number of computing nodes. The role of the middle layer structure is to connect the outer and inner layers. The inner layer contains data mining technology, responsible for providing a platform for the execution of data-related activities in the middle layer. Examples of data-

related work include information sharing, privacy protection and the acquisition of knowledge from areas and applications.

Information sharing is the concern of the whole framework, including processing and big data analytics in smart grid. The outer layer of Fig. 5 shows the data fusion technology necessary for the preprocessing of the heterogeneous, uncertain, incomplete and multi-source data. Complex and dynamic data is extracted after the data preprocessing phase. Subsequently, pervasive smart-grid global knowledge can be obtained through local learning and fusion of the model [31]. Of the decision tree, ridge regression and support vector machine algorithms, the decision tree is found to be the most efficient algorithm for managing energy data based on big data. However, when efficiency is the priority, for example in real-time applications, ridge regression is the most effective algorithm of these three algorithms [51].

Learning is a subfield of machine learning that has the potential to solve a range of complex problems within mobile big data analytics, including classification and regression. Mobile big data samples can be modelled using deep learning consisting of neurons and synapses for training mobile big data samples to learn hierarchical features.

The application of deep learning within mobile big data has the advantages of a high level of accuracy, which is a priority in mobile systems, and multimodal deep learning; intrinsic features are generated by deep learning, necessary in mobile big data analytics, and unlabelled mobile data can be learned using deep learning, which reduces the effort required for data labelling. However, the large number of deep model parameters and the large size of mobile big datasets mean that deep learning is slow and computationally expensive [52].

More recently, deep learning has become a common technique in big data analytics, especially in the retrieval of images with a high level of accuracy [53]. Supervised deep learning and unsupervised deep learning are the two types of deep learning discussed in the literature [54]. A battery with a limited capacity requires an energy efficiency of hundreds of giga floating-point operations per second per watt for a mobile embedded system. This can allow mobile embedded systems to achieve both the required portability and performance [55].

5 Energy Consumption in the Application of Computational Intelligent Algorithms in Big Data Analytics

In the design of computing systems, energy efficiency is one of the most significant issues to be considered. However, the termination of Moore's law has imposed a limit on additional improvements to energy efficiency, which is unfortunate. Recently, the use of physical memristors has shown that it is possible to generate a solution for the integrated hardware of artificial neural networks. This can heavily influence energy efficiency and improve performance [10, 55]. The artificial neural network is one

of the more powerful algorithms in computational intelligence and has received unprecedented attention from researchers; it is believed to constitute one of the major breakthroughs in artificial intelligence. Hu et al. [56] used memristors in the design of a power neuromorphic framework for approximating computation with programmability and computational generality. This design was motivated by the theory of artificial neural networks, which shows that multilayer neural networks are universal approximators, and their wide range of applications in signal processing, pattern recognition, computer vision and natural language processing. A neuromorphic architecture for computing and a tolerance for uncertain computing can generate significant performance and gains in energy efficiency. Wang et al. [55] have found that large-scale artificial neural networks constitute one of the most mainstream algorithms in big data analytics. Two phases are involved in the processing of big data using large-scale artificial neural networks: a training phase and an operational phase. The training phase in big data processing requires a very high amount of computing power and energy efficiency, and this is one of the primary considerations in the operational phase. For example, a ~100 MB training dataset is needed with >100 TOP computation capability, ~40 GB/s IO and SRAM data bandwidth. A 3.4 GHz CPU therefore requires >10 h of learning time for ~100 K input-vector datasets; this requires ~1 s for recognition and is far from real-time processing [57].

The use of computational intelligent algorithms in big data analytics requires high bandwidth interconnection networks with low latency and low power consumption, which are essential for data and storage systems [58]. For example, Wang et al. [55] presented a promising ultrahigh energy-efficient implementation by taking advantage of emerging memristor techniques involving the computing power of GPUs for big data analytics. The results showed a high speedup compared with the basic CPU implementation [55]. Big data analytics using artificial neural networks poses the challenge of how to achieve better training within a lower convergence time and with lower energy consumption [55]. Another computational intelligent algorithm related to energy itself is the deep belief network. The deep belief network consists of the stack of a restricted Boltzmann machine. This is based on the model of energy and certain stochastic methods. Binary values are generated by each of the nodes in the restricted Boltzmann machine, between each node, symmetric link (weight) that can have negative and positive numbers exist. The two types of node in a restricted Boltzmann machine are visible and hidden nodes. The state of a restricted Boltzmann machine is associated with the energy of the restricted Boltzmann machine; a higher energy of restricted Boltzmann machine gives a lower probability of node activation [59].

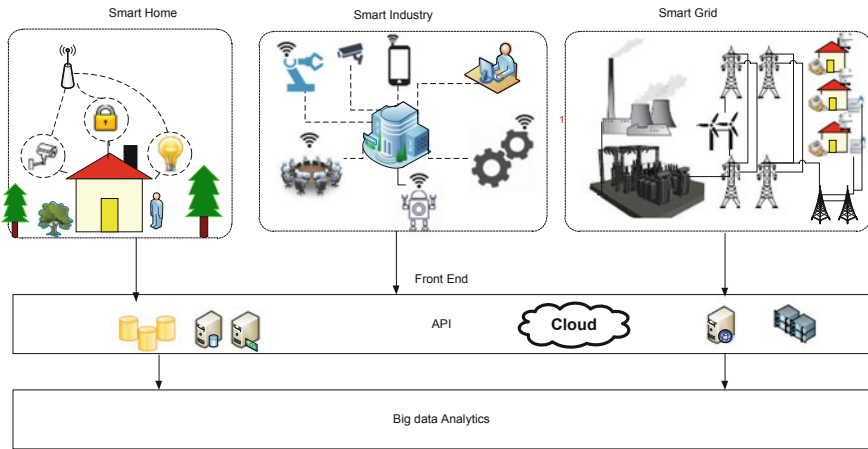


Fig. 6 Large-scale data processing using big data analytics tools such as computational intelligent algorithms

6 A Proposed Framework for Big Data Analytics Using Computational Intelligent Algorithms

Previously, several big data analytics techniques have required an investment in computer hardware and software in order to overcome energy computations. This can result in the downgrading of the performance of the systems used in highly computational environments, which need to process massive amounts of energy data. However, cloud computing and software as a service (SaaS) have now made on-premise solutions unnecessary. In addition, fog computing now allows much of the analytics for these tasks to be moved to the grid edge, to further support the implementation of forecasting and optimisation programmes in real time and at large scales. Two primary big data intelligence applications are load modelling and forecasting, which have been used for energy. These applications are necessary to understand the behaviour of the system in achieving efficient energy management and to enable generic load forecasting [17].

One of the forces driving the adoption of big data analytics is the development of smart grids, although the data generated through these smart grids is increasing in size and difficult to process. Advancements in big data and cloud computing technologies are therefore crucial for a sustainable energy system. Figure 6 shows the processing of large-scale data using big data analytics techniques such as computational intelligent algorithms.

The framework proposes three steps in big data analytics. The first is to collect the required data from sources such as smart industry, smart grid and smart home applications. Subsequently, these data can be stored in a database and servers using cloud storage. A parallel algorithm for optimal power flow based on MapReduce, proposed by Liang et al. [60], is to be applied for power flow calculation in a smart

grid, since this reduces computational complexity. Lastly, the stored data can be processed using computational intelligent algorithms in order to gain insights from the big data.

The application of computational intelligent algorithms consumes a large amount of energy during the training phase as discussed earlier. The proposed framework involves the application of an energy-efficient emerging memristor, which is embedded with an artificial neural network for big data analytics. New and unexpected challenges have been created for the research community, since the current theories and techniques cannot handle big data analytics. Therefore, extension and upgrading of the existing techniques and theories are required to handle big data analytics [4]. As a result, there is no need to propose additional computational intelligent algorithms in the literature (refer to Sect. 2 for a justification); attention should be focused on the modification of the existing ones, since these all have limitations that require improvements in order to allow them to work on big data. The existing computational intelligent algorithms should be modified to handle big data analytics and to require a low energy consumption during the training phase. When processing big data for analytics using a computational intelligent algorithm, a hybrid storage device which combines hard disks and solid-state disks, as proposed by Polato et al. [61], should be used for data storage, as this reduces energy consumption and enhances performance. In addition, the storage of only a proportion of the data in solid-state disks has the potential to significantly save energy and speed up convergence. One typical example of the very high volumes of data generated from smart meters that have the potential to be analysed by computational intelligent algorithms is the case of the 200 TB data generated in Jiangsu, China. Data generation is increasing by 90 GB each day. These data were generated by over 1.81 million acquisition terminals, 1.54 million concentrated meter-reading terminals and 38 million smart meters [62]. Modern data centres should give priority to massive parallel processing in order to enhance computing speeds and reduce energy consumption. In turn, large amounts of data can be moved between the various virtual machines [63]. This framework suggests the use of the following techniques to cushion the high cost of energy consumption by distributed system clusters:

Proportioning the Use of Power The power consumed by hardware components can be proportioned to optimise the operational efficiency of a particular server. The proportioning can also be based on a capped budget, within which the server can be underprovided, in order to be more efficient, or oversubscribed, to operate at high load levels in carrying out its task [31, 62, 64].

Performance Improvement High energy consumption can also be optimised by improving the performance of big data analytical platforms. One example of these improvements is the indexing approach, which works with the main analytical processes and ensures that a full scan of the input data is prevented. This indexing helps to restrict the range of input data to be processed to only that required to carry out the task. Another approach in this category is the scheduling of jobs by considering data locality [65].

Iqbal et al. [66] have pointed out the potential application areas of computational intelligent algorithms within big data analytics which are as follows: personalised health services, biometrics and surveillance, transportation, visualisation of data and interpretation, business and governance, sentimental analysis, models for population displacement, effective computation, fault detection and manufacturing.

7 Conclusions

This chapter presents a theoretical perspective of energy issues within big data analytics, as related to computational intelligent algorithms. Theoretical issues of energy consumption related to big data analytics are described based on computational intelligent algorithms. It is found that the high consumption of energy in big data analytics using computational intelligent algorithms occurs mostly during the training phase of big data processing. We propose a big data analytics theoretical framework based on computational intelligent algorithms with the potential for low energy consumption and performance improvement. The theoretical study presented in this chapter may guide researchers to apply computational intelligent algorithms efficiently and effectively in big data analytics, with the possibility of consuming low energy, and improve performance. Future research directions should be focused on the application of a deep belief network in big data analytics, which consists of the stack of a restricted Boltzmann machine.

References

1. IEA (2016, 5 May, 2017) World Energy Outlook. Available: <http://www.iea.org/newsroom/news/2016/november/world-energy-outlook-2016.html>
2. Horn M, Mirzatury M (2013) Mining big data to transform electricity. In: Broadband networks, smart grids and climate change. Springer, Berlin, pp 47–58
3. Shojafar M, Cordeschi N, Amendola D, Baccarelli E (2015) Energy-saving adaptive computing and traffic engineering for real-time-service data centers. In: 2015 IEEE international conference on communication workshop (ICCW), pp 1800–1806
4. Yu S, Wang C, Liu K, Zomaya AY (2016) Editorial for IEEE access special section on theoretical foundations for big data applications: challenges and opportunities. *IEEE Access* 4:5730–5732
5. Salinas S, Chen X, Ji J, Li P (2016) A tutorial on secure outsourcing of large-scale computations for big data. *IEEE Access* 4:1406–1416
6. Wu K, Barker RJ, Kim MA, Ross KA (2013) Navigating big data with high-throughput, energy-efficient data partitioning. In: ACM SIGARCH computer architecture news, 2013, pp 249–260
7. Li C, Zu Y, Hou B (2016) A feature selection method of power consumption data. In: International conference on computational science and its applications, 2016, pp 547–554
8. Baker T, Al-Dawsari B, Tawfik H, Reid D, Ngoko Y (2015) GreeDi: an energy efficient routing algorithm for big data on cloud. *Ad Hoc Netw* 35:83–96
9. Chiroma H, Abdul-Kareem S, Khan A, Nawi NM, Gital AYU, Shuib L et al (2015) Global warming: predicting OPEC carbon dioxide emissions from petroleum consumption using neural network and hybrid cuckoo search algorithm. *PLoS One* 10:e0136140

10. Li R, Harai H, Asaeda H (2015) An aggregatable name-based routing for energy-efficient data sharing in big data era. *IEEE Access* 3:955–966
11. Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: OSDI'04 Proceedings of the 6th conference on symposium on operating systems design and implementation (Int J Eng Sci Invent). URL: <http://static.googleusercontent.com/media/research.google.com> (diunduh pada 2015-05-10), pp 10–100
12. Kambatla K, Kollias G, Kumar V, Grama A (2014) Trends in big data analytics. *J Parallel Distrib Comput* 74:2561–2573
13. Fernández MR, García AC, Alonso IG, Casanova EZ (2016) Using the Big Data generated by the Smart Home to improve energy efficiency management. *Energy Effi* 9:249–260
14. Abawajy J (2015) Comprehensive analysis of big data variety landscape. *Int J Parallel Emergent Distrib Syst* 30:5–14
15. Wang D, Yu W, Chai T (2015) Guest editorial: special issue on computational intelligence for industrial data processing and analysis. *Neurocomputing* 358–360
16. Cuadra L, Salcedo-Sanz S, Nieto-Borge J, Alexandre E, Rodríguez G (2016) Computational intelligence in wave energy: comprehensive review and case study. *Renew Sustain Energy Rev* 58:1223–1246
17. Hu J, Vasilakos AV (2016) Energy big data analytics and security: challenges and opportunities. *IEEE Trans Smart Grid* 7:2423–2436
18. Engelbrecht AP (2007) Introduction to computational intelligence. In: *Computational intelligence: an introduction*, 2nd edn, pp 1–13
19. Păun G (2005) Bio-inspired computing paradigms (natural computing). In: *Unconventional programming paradigms*, pp 97–97
20. Fister Jr I, Yang X-S, Fister I, Brest J, Fister D (2013) A brief review of nature-inspired algorithms for optimization. arXiv preprint [arXiv:1307.4186](https://arxiv.org/abs/1307.4186)
21. Yang X-S, He X (2016) Nature-inspired optimization algorithms in engineering: overview and applications. In: *Nature-inspired computation in engineering*. Springer, Berlin, pp 1–20
22. Fister Jr I, Mlakar U, Brest J, Fister I (2016) A new population-based nature-inspired algorithm every month: is the current era coming to the end. In: *StuCoSReC: proceedings of the 2016 3rd student computer science research conference*. University of Primorska, Koper, pp 33–37
23. Yang X-S (2014) Cuckoo search and firefly algorithm: overview and analysis. In: *Cuckoo search and firefly algorithm*. Springer, Berlin, pp 1–26
24. Yang X-S, Deb S (2009) Cuckoo search via Lévy flights. In: *World congress on nature & biologically inspired computing*, NaBIC 2009, pp 210–214
25. Yang X-S (2012) Flower pollination algorithm for global optimization. In: *International conference on unconventional computing and natural computation*, 2012, pp 240–249
26. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report-tr06. Erciyes University, Engineering Faculty, Computer Engineering Department
27. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: *Proceedings of the sixth international symposium on micro machine and human science*, MHS'95, pp 39–43
28. Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI. Reprinted in 1998
29. Chen M, Mao S, Liu Y (2014) Big data: a survey. *Mob Netw Appl* 19:171–209
30. Chen CP, Zhang C-Y (2014) Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. *Inf Sci* 275:314–347
31. Gupta A, Gupta S, Ge R, Zong Z (2015) CRUSH: data collection and analysis framework for power capped data intensive computing. In: *2015 sixth international green computing conference and sustainable computing conference (IGSC)*, pp 1–6
32. Yang H-C, Parker DS (2009) Traverse: simplified indexing on large map-reduce-merge clusters. In: *International conference on database systems for advanced applications*, 2009, pp 308–322
33. Jlassi A, Martineau P (2016) Benchmarking Hadoop performance in the cloud-an in depth study of resource management and energy consumption. In: *The 6th international conference on cloud computing and services science*

34. Rabl T, Gómez-Villamor S, Sadoghi M, Muntés-Mulero V, Jacobsen H-A, Mankovskii S (2012) Solving big data challenges for enterprise application performance management. *Proc VLDB Endowment* 5:1724–1735
35. Rong H, Zhang H, Xiao S, Li C, Hu C (2016) Optimizing energy consumption for data centers. *Renew Sustain Energy Rev* 58:674–691
36. Kumar R, Mieritz L (2007) Conceptualizing green IT and data center power and cooling issues. Gartner research paper, 2007
37. Johnson P, Marker T (2009) Data centre energy efficiency product profile. Pitt & Sherry, report to equipment energy efficiency committee (E3) of The Australian Government Department of the Environment, Water, Heritage and the Arts (DEWHA)
38. Karpowicz M, Niewiadomska-Szynkiewicz E, Arabas P, Sikora A (2016) Energy and power efficiency in cloud. In: *Resource management for big data platforms*. Springer, Berlin, pp 97–127
39. Barroso LA, Clidaras J, Hölzle U (2013) The datacenter as a computer: an introduction to the design of warehouse-scale machines. *Synth Lect Comput Archit* 8:1–154
40. Lefurgy C, Rajamani K, Rawson F, Felter W, Kistler M, Keller TW (2003) Energy management for commercial servers. *Computer* 36:39–48
41. Mastelic T, Oleksiak A, Claussen H, Brandic I, Pierson J-M, Vasilakos AV (2015) Cloud computing: survey on energy efficiency. *ACM Comput Surv (CSUR)* 47:33
42. Wang L, Khan SU (2013) Review of performance metrics for green data centers: a taxonomy study. *J Supercomput* 63:639–656
43. Dongarra J, Beckman P, Moore T, Aerts P, Aloisio G, Andre J-C et al (2011) The international exascale software project roadmap. *Int J High Perform Comput Appl* 25:3–60
44. Khalifa S, Elshater Y, Sundaravathan K, Bhat A, Martin P, Imam F et al (2016) The six pillars for building big data analytics ecosystems. *ACM Comput Surv (CSUR)* 49:33
45. Cheng S, Liu B, Shi Y, Jin Y, Li B (2016) Evolutionary computation and big data: key challenges and future directions. In: *International conference on data mining and big data*, pp 3–14
46. Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage* 35:137–144
47. Hashem IAT, Chang V, Anuar NB, Adewole K, Yaqoob I, Gani A et al (2016) The role of big data in smart city. *Int J Inf Manage* 36:748–758
48. Chiroma H, Abdul-Kareem S, Abubakar A (2014) A framework for selecting the optimal technique suitable for application in a data mining task. In: *Future information technology*. Springer, Berlin, pp 163–169
49. Jiang H, Wang K, Wang Y, Gao M, Zhang Y (2016) Energy big data: a survey. *IEEE Access* 4:3844–3861
50. Hu H, Wen Y, Chua T-S, Li X (2014) Toward scalable systems for big data analytics: a technology tutorial. *IEEE Access* 2:652–687
51. Kang D, Kim S, Lee T, Hwang J, Lee S, Jang S et al (2016) Energy information analysis using data algorithms based on big data platform. In: *High performance computing and communications; IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS), 2016 IEEE 18th international conference on*, pp 1530–1531
52. Alsheikh MA, Niyato D, Lin S, Tan H-P, Han Z (2016) Mobile big data analytics using deep learning and apache spark. *IEEE Network* 30:22–29
53. Lee H, Grosse R, Ranganath R, Ng AY (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *Proceedings of the 26th annual international conference on machine learning*, pp 609–616
54. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18:1527–1554
55. Wang Y, Li B, Luo R, Chen Y, Xu N, Yang H (2014) Energy efficient neural networks for big data analytics. In: *Design, automation and test in Europe conference and exhibition (DATE), 2014*, pp 1–2

56. Hu M, Li H, Wu Q, Rose GS (2012) Hardware realization of BSB recall function using memristor crossbar arrays. In: Proceedings of the 49th annual design automation conference, pp 498–503
57. Yoo H, Park S, Bong K, Shin D, Lee J, Choi S (2015) A 1.93 TOPS/W scalable deep learning/inference processor with tetra-parallel MIMD architecture for big data applications. In: IEEE international solid-state circuits conference, pp 80–81
58. Mehdipour F, Noori H, Javadi B (2016) Chapter two-energy-efficient big data analytics in datacenters. *Adv Comput* 100:59–101
59. Park S-W, Park J, Bong K, Shin D, Lee J, Choi S et al (2015) An energy-efficient and scalable deep learning/inference processor with tetra-parallel MIMD architecture for big data applications. *IEEE Trans Biomed Circuits Syst* 9:838–848
60. Liang B, Jin S, Tang W, Sheng W, Liu K (2016) A parallel algorithm of optimal power flow on Hadoop platform. In: Power and energy engineering conference (APPEEC), 2016 IEEE PES Asia-Pacific, pp 566–570
61. Polato I, Barbosa D, Hindle, Kon F (2016) Hadoop energy consumption reduction with hybrid HDFS. In: Proceedings of the 31st annual ACM symposium on applied computing, pp 406–411
62. Nan Z, Hanyong H, Haiyan Z (2016) Efficient stereo index technology for fast combination query of electric power big data. In: 2016 IEEE international conference on computer communication and the internet (ICCCI), pp 329–333
63. Baccarelli E, Cordeschi N, Mei A, Panella M, Shojafar M, Stefa J (2016) Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study. *IEEE Network* 30:54–61
64. Zhu N, Rao L, Liu X, Liu J, Guan H (2011) Taming power peaks in mapreduce clusters. In: ACM SIGCOMM computer communication review, pp 416–417
65. Lee S, Jo J-Y, Kim Y (2016) Performance improvement of mapreduce process by promoting deep data locality. In: 2016 IEEE international conference on data science and advanced analytics (DSAA), pp 292–301
66. Iqbal R, Doctor F, More B, Mahmud S, Yousuf U (2016) Big data analytics: computational intelligence techniques and application areas. *Int J Inf Manage*