IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 4, 2020, accepted January 15, 2020, date of publication January 24, 2020, date of current version February 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969204

Adaptive Fuzzy Map Approach for Accruing Velocity of Big Data Relies on Fireflies Algorithm for Decentralized Decision Making

AYSH M. ALHROOB[1], WAEL JUMAH ALZYADAT[2], IKHLAS HASSAN ALMUKAHEL[1], AND GHAITH M. JARADAT[3]

[1]Department of Software Engineering, Faculty of Information Technology, Isra University, Amman 11622, Jordan
[2]Department of Software Engineering, Al-Zaytoonah University of Jordan, Amman 11191, Jordan
[3]Department of Computer Science, Faculty of Computer Science and Information Technology, Jerash University, Jerash 26150-311, Jordan

Corresponding author: Aysh M. Alhroob (aysh@iu.edu.jo)

ABSTRACT Velocity and volume are two important factors that affect the accuracy of streaming data during the transfer process in Big data applications. This paper presents an Adaptive Fuzzy Map Approach that Relies on Fireflies Algorithm for Accruing Velocity of Big Data and Decentralized Decision Making. A key advantage of the Firefly algorithm is the providing of a small number of iterations comparing to the other methods, which minimize the execution time. Furthermore, the Firefly algorithm is significant to the fuzzy logic system to get its inputs. In addition to the Firefly algorithm, Kalman filter is used to scale the distances of Big data datasets, where it generates output by assigning the match and mismatch. This work used a real dataset to extract variables and values through fuzzification function and be able to coexist as categorical data. After 10 dependent runs that are dealing with certain parameters to be available on aspects of velocity and volume of Big data existing in two parameters Goal and Dimension, the meaningful aspect scale by minimizes the randomness parameter by approximately 1.6%. The other aspect is decision making that is gained through exploration and exploitation that is covered by attraction base and attraction_min parameters. The evaluation has been made by making a comparison between the proposed Adaptive Fuzzy Map Approach and ANOVA model based on the variables like travelled time, road, speed, and distance, which showed clear enhancement produced by the proposed Adaptive Fuzzy Map Approach in terms of the accruing velocity of Big Data.

INDEX TERMS Big data, firefly, fuzzy-map,velocity, decision making.

I. INTRODUCTION

Big Data has a big impact on most real-life aspects; it involves special challenges known as 5Vs that effectively form the way of processing and analyzing to the way of extracting meaning and value from data. These challenges prompted researchers applying many techniques to improve decisions making accuracy, detection of hidden values, and appropriately removing limitations. High dimensionality and massive volume are additionally main issues that characterize Big Data [1]. Using Artificial intelligence (AI) in Big Dat mimics human rationality and anticipation to decide independently. It bases this on the process of learning to gain the required knowledge about the optimal solution and eliminate many paths without having to try them by data analytics and decision-making. Business intelligence is tightly associated with big data analysis techniques. In this respect, AI is used to describe how the data is processed, collected, and analyzed.

Many AI algorithms have been applied to improve big data analysis and recognize patterns; it has also applied them to prediction, as is the case with the firefly algorithm and the well-known Swarm Inelegance model. Additionally. Fuzzy logic plays an important role in exploring data complexity, reducing and clarifying the ambiguity of decision making and

The associate editor coordinating the review of this manuscript and approving it for publication was Xin Luo.

helps in the visualization of knowledge as per the relations among the massive dataset instance.

AI is used to identify the optimal solutions, by three steps firstly look at data for recognizing patterns; second, accumulate the patterns in memory, finally, apply patterns to make inferences or extrapolations and get relevant decisions and reduce difficulties. When beginning from diverse initial values, it executes gradient descent several times and the solution that gives the smallest training squared error as the most promising one. Nevertheless, meta-heuristic algorithms have different solutions. First one, is the single-solution approach, while the second is called a multiple-solution approach. Even in case, the latter method has no capacity of catching local optimal solution alternatives. Consequently, a different method is required to find local optimal alternatives in the solution to this problem in Big Data optimization [2].

The Firefly Algorithm (FA) carries a decentralized behaviour based on a firefly behaviour effort to catch the optimal solution. It is suitable for multimodal optimization. The FA can be characterized as a stochastic, population-based meta-heuristic algorithm that is motivated by the fireflies flashing and their behaviour of acting as a signal system to interest other fireflies as planned in [3].

The FA comprises the following three rules:

1. Firefly algorithm will be concerned with other fireflies regardless of their sex.
2. For any two blinking fireflies, the brighter one will motivate the less bright to move to its position. Otherwise, it will move randomly.
3. The scene of the aim function will determine the blinking of a firefly.

In addition, the variants of the FA are discrete and multi-objective FA. It mainly applies them clustering, continuous optimization, feature selection, and image processing [3].

This work uses FA with a fuzzy controller to convince exploitation and/or exploration relationships. Meanwhile, it is a challenge in big data analysis to avoid the premature convergence problem and develop the decentralization of decision-making; fuzzy logic is used then to handle the randomization output of map-function value of the MapReduce framework. The proposed approach in this work aims at improving the search behaviour of firefly by using an adaptive method to control parameters as a fuzzy logic controller.

The paper is organized as follows: Section II shows the background and related works; Section III presents the proposed work and experiments; section IV deals with the results and discussion and Section V contains the conclusions.

II. BACKGROUND & RELATED WORKS

This section focuses on swarm intelligence algorithms that are used to the sample of data, e.g., reduction of the dimensionality of the big data was developed to handle the big data. It also discusses the techniques of processing big data. The main concern is the velocity accrual for predicting valued decentralized decisions. There is an endeavour to improve the efficacy of big data analysis by obtaining an optimization FA for data analysis. The Fuzzy map approach was utilized for two interconnected reasons. First, most of the data set attributes are irrelevant, resulting in much variance in predictive modelling, and consequently leading to decentralized and increased predictor errors. Second, the variety of Big Data is effectively represented through the analyses of unstructured and semi-structured data. The optimization algorithms and pre-processing analysis seek to acquire meaningfully results as a major process that could improve data quality.

Big data is referred to as the way of describing very large and complex datasets that require advanced data management and visualization techniques. The big data is defined by a common framework known as the five Vs (qualities): volume, variety, variability, value, and velocity. Volume states to the size of the dataset, variety refers to the wide range of data types, veracity refers to the inconsistencies in the data, value determines the benefit of big data, and velocity refers to the amount of data added continuously [4]. In addition, there is one more dimension of the big data known as veracity, which represents the uncertain reliability of the gathered information. For more information, please refer to [5]. Therefore, dealing with these five dimensions of big data requires continuous development of advanced analytics tools. These considerations have motivated us to develop our proposed adaptive fuzzy map approach with the FA in this research work.

In this research, the main concern pertains to the big data velocity. It is known that the velocity determines the speed at which data is being generated, produced, created, or refreshed as stated in [6] and [7], the velocity of data flow has also increased the demand for real-time analytics.

Several distinct research and survey papers provided comprehensive discussions on the challenges of managing big data, approaches, tools, limitations, data processing, and analysis technologies such as [8]; and Networking for Big Data [9]. Velocity problems appear in the form of many issues such as the data structural complexity, lack of consistency, completeness, and multiple dimensionalities. These issues are transformed as Big Data optimization research areas due to handling the high dimensional data; the multiple objectives to a large scale are also important to solve big data problems [10]. AI applies to use the Big Data features and characteristics to enhance the data or rules derived from such data; the result is an optimization of or reliance on machine learning.

One of the widely known and effective techniques of AI is Meta-heuristic. For example, includes genetic algorithms and swarm intelligence algorithms, most of which are inspired by nature. In the last three decades, they have been intensively applied to data in supervised and unsupervised ways; they have also been recently applied to the combinatorial optimization problems, and big data management and analytics.

The key advantage of Meta-heuristics is that the search process balances two major components: exploitation and

exploration [11]. The exploration signifies the process of determining the diverse quality performance in the search space, whereas exploitation focusing on seeking new territory candidate within a search area towards best solutions, thus exploiting the information found so far.

Swarm intelligence is one of the AI disciplines that is inspired by the collective behaviour of social swarms of ants, bees, and flocks of birds. Basically, these swarms consist of unsophisticated individuals, but they exhibit coordinated behaviour that directs the swarms to their desired destinations. This results in the self-organizing behaviour of the whole system; in other words, it results in a multi-agent system [12].

Ibrahim and Simone in [13] proposed a parallel particle swarm optimization clustering algorithm with the MapReduce methodology to analyze the data of an intrusion detection system more effectively in a reasonable amount of time. They proposed a system that processes large datasets on commodity hardware of a real intrusion dataset. Their experimental results demonstrated that the proposed system scales very well with increasing size of the datasets. For example, 80% and 100% of the datasets were determined for training because they are better accuracy-wise. Also, it is compared closely to the linear speedup by improving intrusion detection and false alarm rates. The most important issue in Swarm intelligence performance is the capability to explore and exploit the search areas in order to have a high probability, find out optimal solutions, and focus the search around an optimal solution in order to refine a candidate solution. To attain such a goal, Swarm intelligence must maintain a good balance between the two conflicting objectives of exploration and exploitation [14].

The FA is inspired by the flashing pattern or behaviour of fireflies. It is one of the most recent and important tools of swarm intelligence that have been applied to many optimization problems and other areas such as machine learning. It was first introduced in [3]. As mentioned in [14], the FA belongs to the stochastic (metaheuristic) algorithms, which use a kind of randomization by searching for a set of solutions. In a metaheuristic, the search process of the algorithm is influenced by a certain trade-off between randomization and local search. Hence, it concentrates on the FA on the generation of new solutions within a search space. On the one hand, it selects the best solution for endurance. Alternatively, the randomization empowers the search method to escape the solution being stuck into local optima. The local search expands a nominee solution until expansions are detected. In addition, the FA is a population-based metaheuristic. The algorithms have the following leads when compared to the single-point search algorithms [15]:

- Construction blocks are merged from diverse solutions over the crossover.
- Aiming a search over relies on the crossover and means that if individually parents segment the same value of a variable. Furthermore, the offspring will also have the equivalent value of this variable.
- Low-pass cleaning disregards interruptions within the landscape.
- Hedging in contradiction of bad luck in the original positions or decisions it makes.
- Parameter alteration is the algorithm's chance to learn worthy parameter values to balance exploration (diversification) compared to exploitation (intensification).

The reason behind selecting the FA for the big data domain is based on what was stated by [16], where the FA has two main gains over other algorithms: (i) subdivision automatically and (ii) the capability to deal with multiple local optima. First, FA is based on attraction and attractiveness decreases step by step with distance. This leads to the point that the total population can be automatically segmented into subgroups, and every group can swarm around each local optimum. Among all these local optima, the best solution can be found. Second, this subdivision allows the fireflies to be able to find all optima concurrently if the population size is appropriately higher than the number of local optima. Scientifically, the reduction factor of the FA ($1/\sqrt{\gamma}$) joysticks the average distance of a group of fireflies that can be seen by adjacent groups. Therefore, a whole population can be subdivided into subgroups with a given average distance. This automatic subdivision ability makes the FA particularly suitable for highly nonlinear, multimodal optimization problems. In addition, the parameters in FA can be tuned to control the randomness as iterations proceed, so that convergence can also be sped up by tuning these parameters. These above-mentioned advantages make the FA flexible to deal with continuous problems, such as clustering and classifications when dealing with big data, and combinatorial optimization as well.

In addition, [16] stated that the FA is relatively simple in terms of complexity and implementation. The FA has two inner loops in the population (n) and one outer loop for the iteration (t). So the complexity at the worst case is $O(n^2t)$, where the computation cost is relatively inexpensive because the algorithm complexity is linear in terms of iteration. Typically, the main computational cost will be in the evaluations of objective functions, especially for all optimization problems but for the big data this might not be the case. As recommended by [16] we have considered using only one inner loop in our proposed FA for large population size by classification the brightness or attractiveness of all fireflies with sorting algorithms. At that time, the algorithm complexity of the FA will be $O(ntlog(n))$.

The light strength of Firefly is a comparative measure of the value and quality of the optimal solution; it inspires fireflies that have brighter light and decreased light intensity. It can be emulated well by following equation [17]:

$$\text{Attractiveness } (10, \gamma, \text{d}) = \frac{\text{L}_0}{1 + \gamma \text{d}^2} \tag{1}$$

While: γ is the reduction coefficient L_0 is the light intensity at the source and (d) is the distance.

Furthermore, the distance scale in the FA provides information about the diversity between two fireflies i as well as j, at positions x^i and x^j, the distance function can be expressed as follows:

$$\text{Distance}\left(x^i - x^j\right) = \sqrt{\sum_{k=1}^{n}\left(x_k^i - x_k^j\right)2} \qquad (2)$$

where in equation (2) x^i and x^j represents k^{th} component of spatial coordinate x^i of i^{th} firefly. However, the movement describes the step of the firefly in following the way towards the better solution; it determines the maximum radius of the random step. This relation is important for the exploration-exploitation to maintain the balance for the most efficient search [18].

FA has some limitations, affecting the performance of the parameters constant and fixed-on run time. FA does not hold any history of the optimal solution for each firefly and moves in spite of it; this may lead to missing the optimal solution [19]. Using the Fuzzy Logic Controller indicates the major drawback of FA, which almost converges to a local minimum or a saddle point.

According to [6], the authors used the fuzzy logic controller as a parameter controller in the FA to tune parameters to keep a balance between exploration and exploitation. This method helps in addressing big data analysis issues and improving quality towards reaching an optimal pattern and decision-making.

Previous work has focused on a specific characteristic of big data such as velocity and volume. According to [20], ''Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.'' Based on the nature of data, the Big Data volume is the main dimensions.

Authors in [20] addressed the challenge of big data with respect to velocity, showing it by combining big data and AI techniques. Hence, they have proposed an approach based on Fuzzy Logic and MapReduce to create an optimal dataset. A fuzzy map is also used for the accruing velocity of big data. Their approach handles the velocity and assures the performance of big data analysis by reducing the processing time. The proposed approach proved effective, and it can be applied to other real-world applications.

The velocity is the speed of data creation (i.e. real-time data) this issues illustrated and analysis by techniques as Machine Learning to Predict the disease risk and Natural Language Processing for Improvement the efficiency of care and controlling costs.

It is significant to notice that not all big data challenges can be handled using the Fuzzy-MapReduce approach. It is necessary that the problem is divided into key and value pairs, and that all key-values are divided according to the rule base extracted from the fuzzy controller. However, it is also important to decide how many new points will be generated through each map operation. It is possible to use the best n points to create new values [21]. Therefore, we are motivated to develop a hybrid approach, which analyzes big data using the FA and processes, using an adaptive fuzzy map approach that focuses on velocity accrual. Therefore, we are motivated to develop a hybrid approach to overcome the downside of working with Fuzzy MapReduce approach. The reason being that it is not efficient to determine the rule according to dataset attributes without depending on the parameter of search space and candidates solution; we focused, therefore, on firefly heuristic algorithm to accrue velocity of big data.

FA and Fuzzy approach are used in several domains to obtain an enhanced outcome. In the network domain, Lalwani *et al.* [22] proposed a new algorithm to solve a related problem using FA. Consequently, the Routing algorithm is developed with a novel fitness function based on residual energy, node degree, and distance. Thus, it finds out the complete routing solution for multi-hope communication between the Cluster Head and Base Station.

In the image-processing domain, authors in [23] proposed a version of the glistenings Detection technique. Fuzzy-based FA is used in feature selection, applied the concept of fuzzy-entropy to determine the membership of features in order to select good sets of the significant features that make the most of the classification performance in glistenings Detection. Their proposed method was then compared to the feature selection methods and the standard FA using basic classifier k-nearest neighbour without selection. Results showed that the Matthews correlation coefficient and the diagnostic odds ratio value increase after feature selection when using the FA and fuzzy entropy. A small portion of the features set has also decreased the classification time during the testing phase.

Authors in [20] used the Fuzzy logic controller, combined with MapReduce frameworks, to handle the vehicle analysis by comparing the driving data from the new outcome vehicle trajectory. The proposed approach was evaluated via the amount of raw data from the original resource with dataset after processing the approach by using Analysis of variance (*ANOVA*) to estimate and analyze the differences.

ANOVA is a statistical method used to verify two or more organizations differ considerably. The difference before and after using the said approach is a positive impact in several stages of the volume of datasets, variances, and P-value, demonstrating a significant meaning and contribution to the two aspects of accuracy and performance. Nevertheless, one of the main limitations involved decreasing the distance among indexing datasets to enhance dynamic classified data. The proposed method in this work suggests using the FA to overcome the limitation in previous work [20].

As Authors state FA are among the most powerful algorithms for optimization. One of our motivations in the present study is to compare the performance of Fuzzy logic controller, combined with MapReduce for a real-world dataset with firefly combined algorithm.

III. PROPOSED METHODOLOGY

In this section, an adaptive FA is represented, combined with fuzzy map approach, to accrue velocity of big data as well

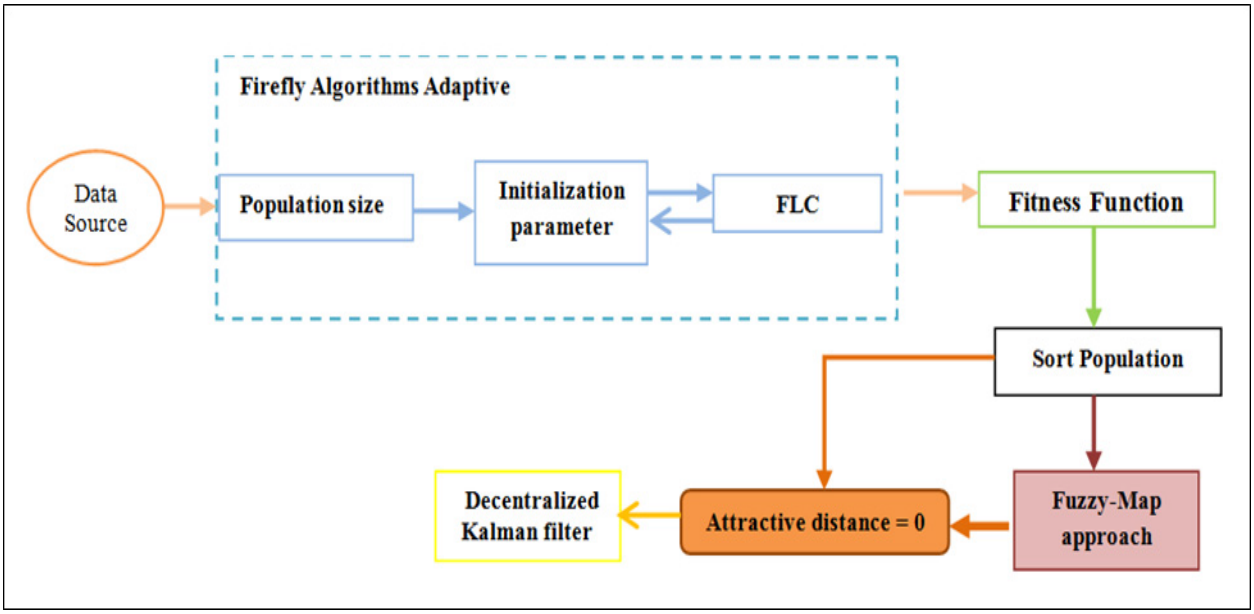

FIGURE 1. Adaptive FA combined with fuzzy map approach for accruing velocity of big data.

as value by finding the Distance among solutions to reach the optimal decisions in a suitable time. Seven components are applied to attain the best possible decentralization decision. The first component pertains to obtaining input data as available. Once the input gets a hold, the second component launches the FA to determine the population size and start the heuristic search and control. Meanwhile, the firefly parameters are considered as inputs using a fuzzy logic controller. The third component is responsible for the computing fitness function. Consequently, the fourth component sorts the population and prepares data for the fuzzy map approach. The fifth component represents the relationship between datasets and assures velocity. Thus, if the attractive distance $= 0$, it will move to the *Kalman* filter. Finally, the duty of the *Kalman* filter component is to track the matching percentage between raw (input data) and the new dataset as shown in Figure1.

The components of the Fuzzy-Map approach for accruing velocity of Big Data interact as follows:

1- Component one is data collection and preprocessing, and provides the input data into approach from the source of data is https://www.kaggle.com/zhaopengyun/driving-data/home; the key actors for generate datasets are a vehicle and video camera. In addition, the contextualization of mechanism data is kinematics that request to layout the synchronized network. The data was captured from southbound US 101 and Lanker Shim Boulevard in Los Angeles, CA, eastbound I-80 in Emeryville, CA and Peachtree Street in Atlanta, Georgia used software application named NGSIM.

2- Component two is a firefly adaptive algorithm that uses a fuzzy controller to determine the value of and update the FA while running.

3- Component three is consequently used to calculate the degree of a fitness function from each parent in the population sample.

4- Component four is designed to sort the population according to fitness function; it is an important component to be in parallel with the linguistic term, ensuring that diversity is attained with a minimum fitness function. Meanwhile, the process is repeated for all parents in the population to construct a Fuzzy controller and update the firefly parameters.

5- Component five uses the Fuzzy-Map approach to refine and divide the output from population and attendance. The If-Then rule is used to reduce random variables produced by the MapReduce application to organize the population. The output content of this component aims to replace the random value of the degree of the fuzzy rule mapping; the latter is responsible for producing a result in logic, in light of the Fuzzy sets and membership degrees. In other words, it is the process that maps a Fuzzy set to a crisp set.

6- Component six: As a result, While the iteration does not reach stopping condition which is attractive function value $=$ 0 the approach is not reaching the optimal solution.

7- Component seven is the filtering technique that can evaluate the differences between data sets in terms of the overwhelming size. In this respect, we apply The *Kalman* filter for tracking and assessment due to its ease, optimality, tractability, and robustness to compute the different distance among data sets which, in turn, can determine a match or a mismatch status among old and new datasets.

The existing components from the Adaptive FA were combined with Fuzzy Map Approach to accrue Velocity of Big Data; they are responsible for investigating an accurate decision with big data. Figure 2 shows the interaction between the above components as following steps:

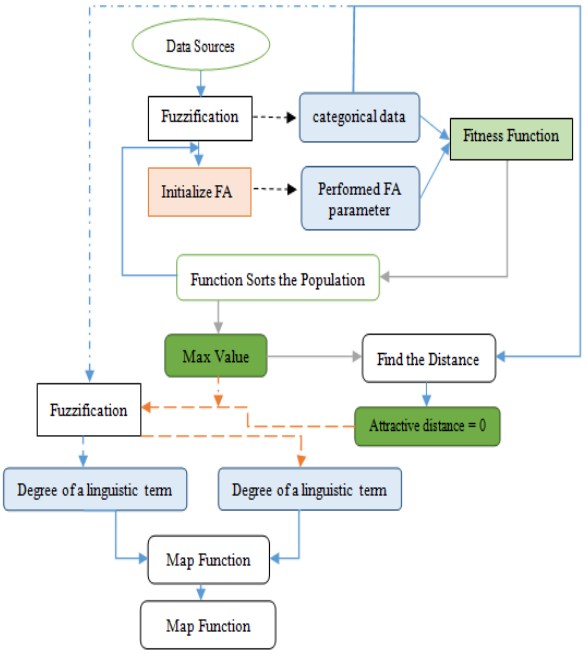

FIGURE 2. FA combined with fuzzy map approach component.

1- Input data collection, from the source and preprocessing it.
2- Applying fuzzification (membership function) also determine the value of firefly parameter.
3- Initialize firefly adaptive algorithm that uses a fuzzy value.
4- Calculate the degree of a fitness function from each parent in the population sample.
5- Sort population according to the fitness function.
6- Applying Fuzzy-Map approach to refine and divide the output from population and attendance.
7- Check stopping condition which is attractive function value $=0$ the approach is not reaching the optimal solution then do fuzzification again.
8- Evaluate the differences by applying the *Kalman* Filter for tracking and determine a match or a mismatch status among old and new datasets.

The Adaptive FA combined with Fuzzy Map Approach for Accruing Velocity of Big Data uses R language, Microsoft Excel 2016 and some R packages are used as shown in Table 1.

The data collection and preprocess fetch the raw dataset; it includes 82 attributes and 17897 records of different types such as the string, numeric, and Boolean. Furthermore, the quintiles description for each attribute is used in various ways to be observed by the mean and standard deviation. Once variables are defined, the next step is defining population size and the Firefly parameter as shown in Table 2.

The first role of the Fuzzy logic controller is to determine the most suitable value for each parameter that is important to investigate new areas in a search space and make use of knowledge acquired by exploration to reach a better decision

TABLE 1. *R* packages use to experiment.

R tool Packages	Source	Exact of use in the proposed approach
Read Rectangular Text Data	https://rdrr.io/cran/readr/	Read source data set
Dplyr	https://cran.r-project.org/web/packages/dplyr/index.html	Data manipulation and processing
Tidyr	https://cran.r-project.org/web/packages/tidyr/index.html	Work with Attributes (column) and Raw (Observations)
metaheuristicOpt	https://cran.r-project.org/web/packages/metaheuristicOpt/index.html	implementation of metaheuristic algorithms
Classification and Regression Training caret:	https://cran.r-project.org/web/packages/caret/index.html	Preprocess data set and training
Hadoop Streaming	https://hadoop.apache.org/docs/r1.2.1/streaming.html	Provides a framework for writing map/reduce
HiveR	https://cran.r-project.org/web/packages/HiveR/index.html	Function map, manager and plots
FuzzyR	https://cran.r-project.org/web/packages/FuzzyR/index.html	Design and simulate fuzzy logic

TABLE 2. Firefly controller parameters.

Parameter	Description
Goal	Goal function (stop.concition)
Dimensions	Number of dimensions of the problem to solve
Params	List of additional parameters
fflies_count	Fireflies population size
Iterations	Iterations count
Coefficients	Calculation coefficients
Absorption	Light absorption over a distance
Randomness	Randomness coefficient
attraction_base	Base attraction
attraction_min	Min attraction
Min	The minimal size of the world
Max	The maximal size of the world

on the search space. By performance, measure holds on after applying firefly with initial value adaptable – the value which is most likely to achieve goal function and fixation. All parameter values are shown in Figure 3.

After adaptive control parameter of FA (FLC), the parameter value the fitness function component is applying calculates the fitness function for each population this component is important to produce the degree, which a

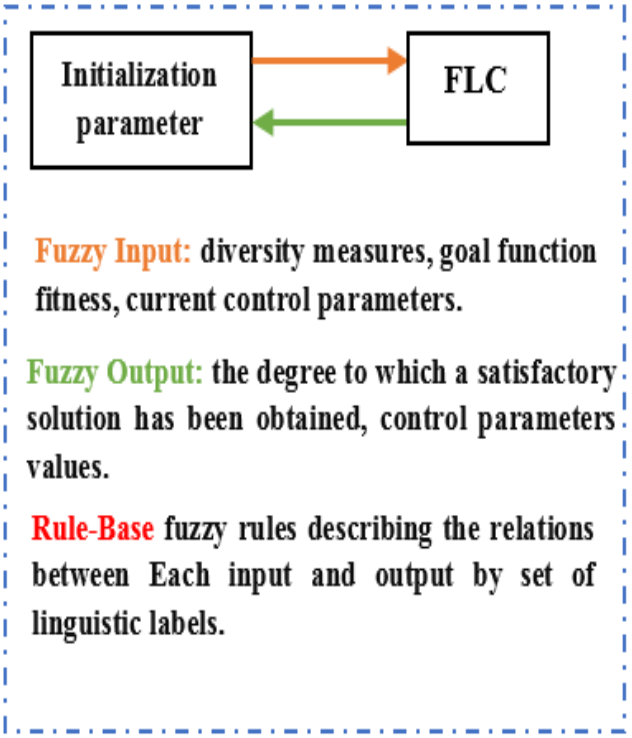

FIGURE 3. Adaptive firefly parameter by FLC.

TABLE 3. Variables consists of fuzzy.

Variables	Values
Vehicle	bad, ok, and perfect
Distance travelled	near, middle, and far
Speed	slow, limit, and fast
Road statues	crowded, normal, and open

satisfactory decision will obtain. Then sorting all population according to fitness function and find the maximum, minimum and average helps in decentralization taxonomy to determine the direction that exploitation and exploration relationship kept throughout the run. Fuzzy map approach first defines rules, which are the links between the ''non-final'' variables, which are the distance travelled, speed, road status, and vehicle, as shown in Table 3.

Dynamically adjust selected most effective rule that extracts by membership function important role in the enhancement and the performance of Big Data in term of velocity.

IV. RESULTS DISCUSSION

The results show evidence and proofs that the Adaptive FA combined with Fuzzy Map Approach plays an important role in dealing with big data characteristic such as value by applying firefly and fuzzy and accruing velocity of big data by fuzzy map approach. First fuzzification function applies to extract a fuzzy membership among all variable. Then instance

TABLE 4. Normalization variables key.

Variables	Values
Vehicle	Bad =1, ok =2, and perfect =3
Distance traveled	Near =4, middle =5, and far =6
Speed	Slow =7, limit =8, and fast =9
Road statues	Crowded=10, normal=11, and open=12

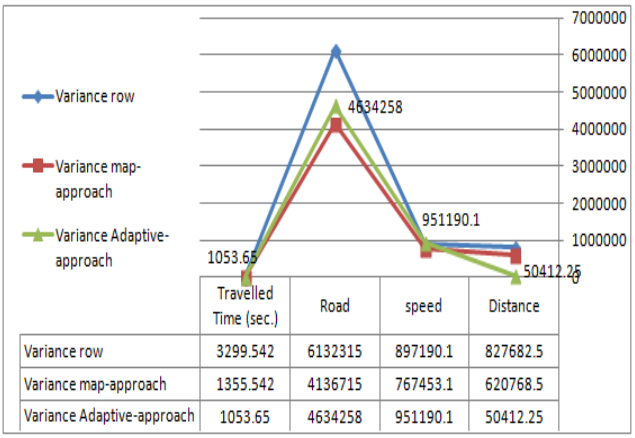

FIGURE 4. ANOVA applied to raw dataset, after fuzzy-map, and adaptive approach.

is converted for categorical data finally normalization applies to smooth the progress of statically analysis as a result shown in Table 4.

Each normalization value is a range of instance value depending on membership function the best solution reach when values of each variable achieve the best Vehicle $=$ 3, $+$perfect Distance $=$ 4travelled Speed $+$ 9Road statues $+$ 12. The result illustrated in Table 5, shows the new firefly controller parameter value by applying the adaptive firefly by Fuzzy controller this value are the average values after 10 dependent run and hold on the result.

Once the controller parameter is assigned. Next component applies fitness functions (goal) and sorting them, this step responsible for decreasing the difficulty of a problem by determining the boundary of research space. Sequentially applying fuzzy map approach to produce the relationship among variables for accruing value and velocity from big data aspect. Figure 4 illustrates the variance among row data, fuzzy-map approach, and adaptive approach.

Distance in raw dataset variance is 827682.5 in adaptive approach variance 50412.25, which means the adaptive approach reduces the distance as a result of finding the shortest path. Speed raw dataset variance is 897190.1, an adaptive approach variance 951190.1, which means the new dataset reduces time better than the raw. Road raw dataset variance is 6132315; in new approach variance 4634258 that evidence the new dataset chose the shortest road as a result of firefly adaptive via Fuzzy controller. Time-travelling from raw dataset variance is 3299.542, in new dataset variance

TABLE 5. Difference between the two datasets.

Parameter	Initial Value	Fuzzy Adaptive	Effective
goal	$f(x) = \sum_{n=1}^{4} (\|xi\|) - 2$ The global maxima is located at x = f (3, 4,9,12).	Rule base depends on two variables x1 and x2 We can express the following as a math model: F(x) = max ($\|x_1 + 2x_2 - 6\|$, $\|2x1 + x2 + 7\|$) subject to $-100 \le xi \le 100$. The global maxima are located at x = f (3, 5).	Reduce dimensionality and deal with decentralization by more effective way
dimensions	4^3 = 64	2^2 = 4	
fflies_count	10	18	Reduce time and increase exploration
iterations	1000	1234	Check that not trapped in local optima
coefficients	0.5	0.8	Achieve converge and exploration
absorption	0.7	1	
randomness	0.3	0.05	Exploitation
attraction base	0.6	0.9	Increase exploration
attraction_min	0.4	0.09	Exploitation
Min	10000	No change	-
Max	17897	No change	-

TABLE 6. Results achieved by the standard FA and AFMA, where the best results are shown in BOLD.

	FA				AFMA			
	Worst	Best	Mean	Std dev	Worst	Best	Mean	Std dev
Traveled time	7.534 6.5643	5.434	6.6704	1.834	6.3544	**1.351**	**2.0244**	2.8544
Distance	3.865	2.0811	**4.1902**	**1.422**	9.0811	4.5416	5.8312	4.0112
Road	9.562	3.2361	2.4305	4.854	1.0201	**3.8500**	**2.0101**	5.9200
Speed		5.607	8.3501	3.1601	14.560 7	**12.0909**	**9.3007**	8.8508
Compute - time	9.456E-9	6.354E-4	2.69E-8	6.210E+7	6.354E-4	**1.635E-1**	**2.024E-4**	2.85E-44

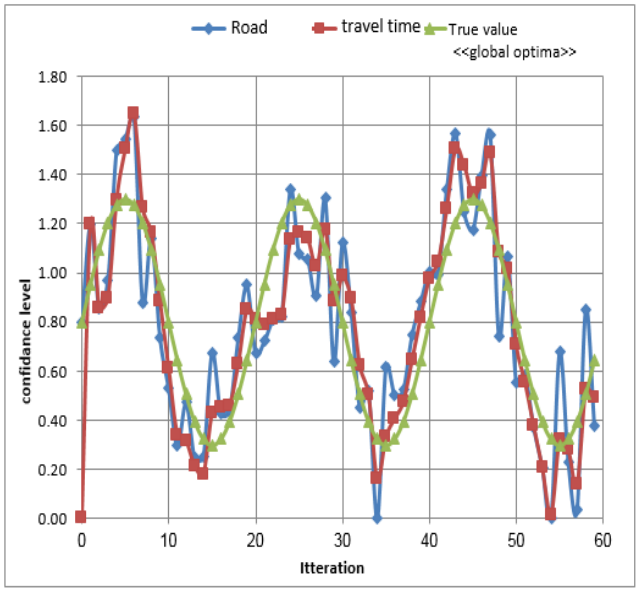

FIGURE 5. Road and travel time.

1053.65 that means the new dataset reach distance within minimum travelling time.

As shown in Table 5, a new adaptive FA parameter using Fuzzy-map approach can converge to the global optimum on using a new dataset parameter. AFMA can adjust the value and find global optimum, it usually converges to some suboptimal solutions Table 6 shows the computational results of FA combined with FLC on the dataset. It can be seen to find promising optimal value on distance, travel time, road, and speed parameters. As well shown best performance from computational time aspect using firefly flash from *abd* package in *R*.

Analysis of the variance for optimal decision with the global optima by applying scalar *Kalman* filter, comparing the best value of the variable with *Kalman's* before and after applying the approach. to determine whether there a significant and satisfactory degree for the investigated result for the four variables independent unrelated filtering and dependent filtering groups as a true rule to reach confidence as well as sense decision applying ideal driving behaviour. Figures 4 represents the adaptive data set that acquired by applying adaptive firefly approach variance is more efficient in reducing the cost of the process in autonomic vehicles this reduction is due to the converge relation kept in data after applying to map the fuel used and distance also the travelling time will reduce.

After applying fuzzy-map, the optimal solution for road and travel time estimated by discrete-time and observations were taken at a frequency of 60 iterations (see Figure 5). The best confidence level of Road = 1.65 that's mean the best road

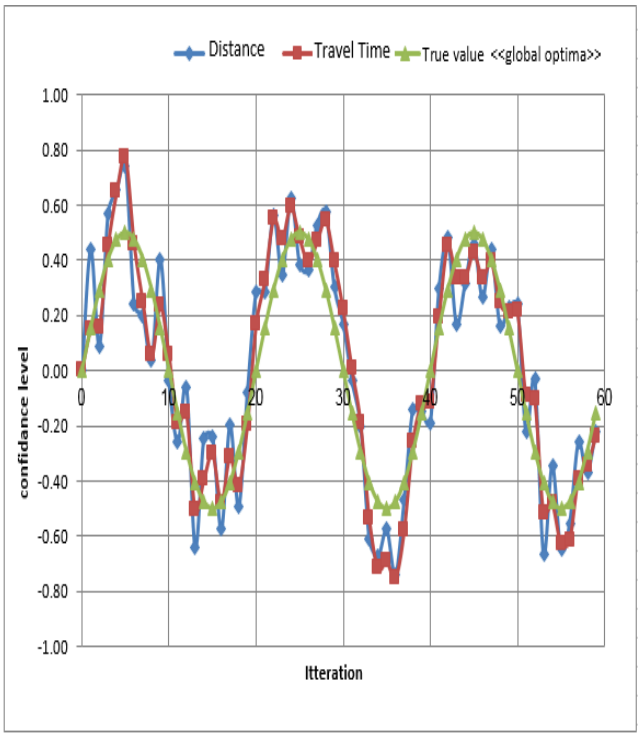

FIGURE 6. Distance and travel time predictable.

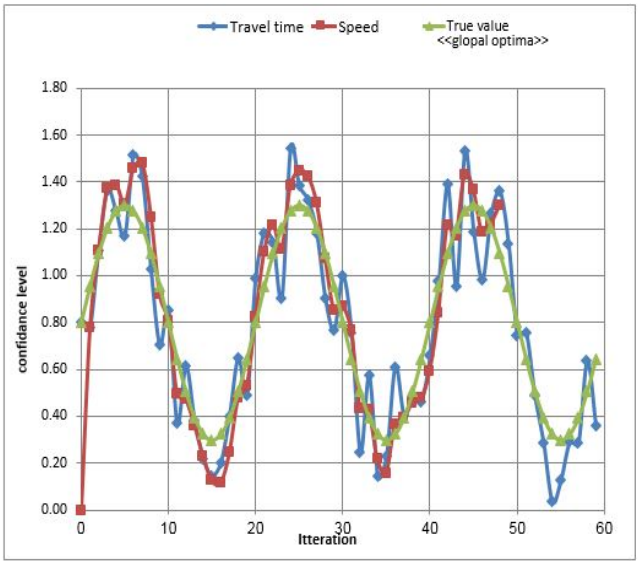

FIGURE 7. Speed and travel time predictable.

status achieve in 1.65% travel time that's mean 1.the optimal converge between achieving as global optima regarding to balance between Road and travel time.

The optimal solution according to distance and travel time predictable after normalization as reduce distance and travel time as possible.by the regularity of 60 iterations, the best confidence level of distance = -.60 that's mean the longest distance can achieve = 2896411.25 meter in 0.8 travel time that's mean 1 as shown in Figure 6. The optimal converge between achieving global optima regarding the balance

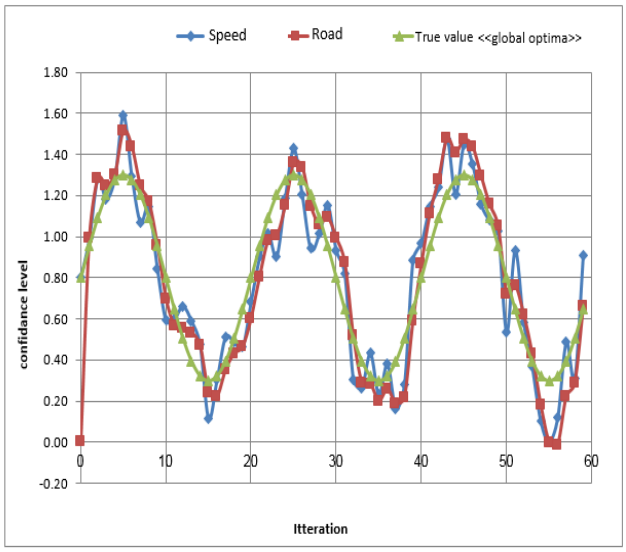

FIGURE 8. Kalman filter mong variable and truth value.

between distance and travel time. Figure 7 illustrates that the Speed value is limited according to the speed scale that means when the optimal distance or road status achieve within the specific time value mainly = 1.4% to 1.65% travel time.

According to Figure 8 after apply scalar *Kalman*filter between Speed and road, the optimal speed achieves when the road status is optimal this means that the travel time depends on both speed and road statues even if the distance is not optimal.

V. CONCLUSION

This research presents an approach dealing with the characteristics of big data (velocity and volume), by using a metaheuristic called firefly optimization algorithm. The research aimed at improving the search behaviour of firefly by using an adaptive method to control parameters as a fuzzy logic controller. The approach proof improving the performance of big data analysis by reducing the processing time towards an optimal solution. In addition, reduce the dimensionality by applying Firefly headed for decentralization decision making. The main improvement of the adaptive approach is to define the parameter on run time and extract rule base to build knowledge-rule based the search space increases exponentially by the number of variables. Therefore, deciding the effective value must depend on the weight and correlation among values.

FA has made a positive impact on a fuzzy map approach, the significances are ahead of Big Data characteristics, velocity characteristic face challenge of behaviour source data, which reflect accuracy and meaningful data. On other hands, the volume characteristic compound to exploitation and exploration, both should be put together to optimize the indexing data and determine the relationship via the fuzzy controller. Fuzzy controller, in short, is reducing the processing time and reduce the dimensionality. Our experimental results that show the proposed approach, deciding the

effective value depend on the weight and correlation among other values throughput define the parameter on run time and extract rule base. Our findings showed the importance of the combination between an optimization approach with the fuzzy logic to overcome computational time complexity and controlling parameters. Also, in this study, the FA has been applied for the first on big data, in terms of data type manipulation and controlling parameters of the learning method, namely here, the fuzzy logic.

Although, our approach is somewhat limited to the content of the handled data or the application. However, it could be extended to different applications. In future works, we would apply our approach to different fields of big data. Such fields: big data in business, marketing, and sentiment analysis.

REFERENCES

[1] S. Ramírez-Gallego, A. Fernández, S. García, M. Chen, and F. Herrera, "Big Data: Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce," *Inf. Fusion*, vol. 42, pp. 51–61, Jul. 2018.
[2] M. Kilinc and J. M. Caicedo, "Finding plausible optimal solutions in engineering problems using an adaptive genetic algorithm," *Adv. Civil Eng.*, vol. 2019, pp. 1–9, Feb. 2019.
[3] X.-S. Yang and X.-S. He, "Why the firefly algorithm works?" in *Nature-Inspired Algorithms and Applied Optimization*. Cham, Switzerland: Springer, 2018, pp. 245–259.
[4] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in *Proc. 46th Hawaii Int. Conf. Syst. Sci.*, Piscataway, NJ, USA, 2013, pp. 995–1004.
[5] T. Usha Rani, C. Sindhu Priyanka, and B. S. S. Monica, "A dynamic data classification techniques and tools for big data," *J. Phys., Conf. Ser.*, vol. 1228, May 2019, Art. no. 012043.
[6] M. Bidar, S. Sadaoui, M. Mouhoub, and M. Bidar, "Enhanced firefly algorithm using fuzzy parameter tuner," *Comput. Inf. Sci.*, vol. 11, no. 1, pp. 26–51, 2018.
[7] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015.
[8] M. Adiba, J. C. Castrejon-Castillo, J. A. E. Oviedo, G. Vargas-Solar, and J.-L. Zechinelli-Martini, "Big data management challenges, approaches, tools and their limitations," in *Networking for Big Data*. London, U.K.: Chapman & Hall, 2016, pp. 1–22.
[9] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Comput. Hum. Behav.*, vol. 101, pp. 417–428, Dec. 2018.
[10] S. Amini, I. Gerostathopoulos, and C. Prehofer, "Big data analytics architecture for real-time traffic control," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Naples, Italy, Jun. 2017, pp. 710–715.
[11] M. Crepinšek, S. H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey," *ACM Comput. Surv.*, vol. 45, no. 3, p. 35, 2013.
[12] A. Girdhar, "Swarm Intelligence and flocking behavior," *Int. J. Comput. Appl.*, pp. 0975–8887, 2015.
[13] A. Ibrahim and A. L. Simone, "Apreduce intrusion detection system based on a particle swarm optimization clustering algorithm," in *Proc. IEEE Congr. Evol. Comput.*, Cancún, Mexico, 2013.
[14] A. Cano, C. García-Martínez, and S. Ventura, "Extremely high-dimensional optimization with MapReduce: Scaling functions and algorithm," *Inf. Sci.*, vols. 415–416, pp. 110–127, Nov. 2017.
[15] A. Prugel-Bennett, "Benefits of a population: Five mechanisms that advantage population-based algorithms," *IEEE Trans. Evol. Comput.*, vol. 14, no. 4, pp. 500–517, Aug. 2010.
[16] X.-S. Yang and X. He, "Firefly algorithm: Recent advances and applications," *Int. J. Swarm Intell.*, vol. 1, no. 1, pp. 36–50, 2013.
[17] S. Łukasik and S. Żak, "Firefly algorithm for continuous constrained optimization tasks," in *Proc. Int. Conf. Comput. Collective Intell.*, Bristol, U.K., 2009, pp. 97–106.
[18] J. Aswini, N. Malarvizhi, and T. Kumanan, "A novel firefly algorithm based load balancing approach for cloud computing," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, no. 5, pp. 91–96, 2019.
[19] K. Murugasamy and K. Murugasamy, "Hybrid clustering using firefly optimization and fuzzy c-means algorithm," *Circuits Syst.*, vol. 7, no. 9, pp. 23–39, 2016.
[20] W. J. Alzyadat, A. AlHroob, I. H. Almukahel, and R. Atan, "Fuzzy map approach for accruing velocity of big data," *COMPUSOFT, Int. J. Adv. Comput. Technol.*, vol. 8, no. 4, pp. 3112–3116, 2019.
[21] C. Jin, C. Vecchiola, and R. Buyya, "MRPGA: An extension of MapReduce for parallelizing genetic algorithms," in *Proc. IEEE 4th Int. Conf. eScience*, Indianapolis, IN, USA, Dec. 2008, pp. 214–221.
[22] P. Lalwani, I. Ganguli, and H. Banka, "FARW: Firefly algorithm for Routing in wireless sensor networks," in *Proc. 3rd Int. Conf. Recent Adv. Inf. Technol. (RAIT)*, Jharkhand, India, 2016, pp. 248–252.
[23] P. Jitpakde, B. Uyyanonvar, and C. Hull, "Feature selection using fuzzy-based firefly algorithm for glistenings detection on intraocular lenses," *J. Comput.*, vol. 30, no. 2, pp. 103–115, 2019.

AYSH M. ALHROOB received the Ph.D. degree from the University of Bradford, U.K., in 2010. He is currently an Associate Professor of software engineering with Isra University, Jordan. He joined Isra University as an Assistant Professor with the Software Engineering Department, Faculty of Information Technology. He has a number of published articles in various computer science and software engineering topics in international journals and conferences. In addition, he has published his first book in software testing, in 2010.

WAEL JUMAH ALZYADAT is currently an Assistant Professor of software engineering. He also works with Al-Zaytoonah University, Jordan. His research area encompasses the area of software analysis, intelligence systems, streaming data, and big data. Moreover, he established more than 20 published articles and achieved two copyrights.

IKHLAS HASSAN ALMUKAHEL received the bachelor's degree in computer software engineering from Science and Technology University, Sana'a, Yemen, and the master's degree in software engineering from Isra University, Amman, Jordan. She is currently a Teaching Assistant with the Software Engineering Department, Faculty of Information Technology, Isra University. She is also the coordinator of the Quality Department. Her research interests are in the areas of software modeling, artificial intelligence (AI), big data, and analytics data. She serves on Steering Committee for big data coherences with KDD.

GHAITH M. JARADAT received the bachelor's degree from the Computer Science Department, Jerash University, Jordan, in 2004, the master's degree from the Intelligent Systems Department, Utara University, Malaysia, in 2007, and the Ph.D. degree in intelligent research algorithms-computer science from the National University of Malaysia, in April 2012. He is currently an Associate Professor with the Department of Computer Science, Faculty of Computer Science and Information Technology, Jerash University. He has published a number of high-quality research articles in international journals and conferences. His research interests are mainly directed to metaheuristics and combinatorial optimization problems, including course and exam timetabling, vehicle routing, travelling salesman, knapsack, and nurse rostering problems.