International Journal of Computer science engineering Techniques-– Volume 1 Issue3, Mar - Apr 2016

RESEARCH ARTICLE OPEN ACCESS

Efficient Methods for Data Science using Various Population-Based Algorithms

Kotadi chinnaiah[1], Akash saxena [2]

[1](Rescarch scholor, depaptment cse, sunrise university, alwar, rajasthan, india)
[2](Rescarch superviser, depaptment cse, sunrise university, alwar, rajasthan, india)

Abstract

This paper discusses the relationship between data science and population-based algorithms, which include swarm intelligence and evolutionary algorithms. We reviewed two categories of literature, which include population-based algorithms solving data analysis problem and utilizing data analysis methods in population-based algorithms. With the exponential increment of data, the data science, or more specifically, the big data analytics has gained increasing attention among researchers. New and more efficient algorithms should be designed to handle this massive data problem. Based on the combination of population-based algorithms and data mining techniques, we understand better the insights of data analytics, and design more efficient algorithms to solve real-world big data analytics problems. Also, the weakness and strength of population-based algorithms could be analyzed via the data analytics along the optimization process, a crucial entity in population-based algorithms.

Keywords: Big data analytics, Data analysis, Data science, Evolutionary algorithms, Swarm intelligence, Population-based algorithm

I. Introduction

With the amount of data growing constantly and exponentially, the current data process-ing tasks are beyond the computing ability of traditional computational models. The data science, or more specifically, the big data analytics, has attracted more and more atten-tion among researchers. The data are easily generated and gathered, while the volume of data is increasing very quickly. It exceeds the computational capacity of current systems to validate, analyze, visualize, store, and extract information. To analyze these massive data, there are several kinds of difficulties, such as the large volume of data, dynami-cal changes of data, data noise, etc. New and efficient algorithms should be designed to handle massive data analytics problems.

Swarm intelligence and evolutionary algorithms are two sets of search and optimization techniques [1–3]. To search a problem domain, a swarm intelligence algorithm processes a population of individuals. Different from traditional single-point based algorithms such as hill-climbing algorithms, each swarm intelligence algorithm is a population-based algorithm, which consists of a set of points (population of individuals). Each individ-ual represents a potential solution to the problem being optimized. The population of individuals is expected to have high tendency to move towards better and better solution areas along iteration through cooperation and competition among themselves.

In this paper, we present the analysis of the relationship from data science to population-based algorithms, which include swarm intelligence and evolutionary algo-rithms. Swarm intelligence/evolutionary algorithms could be applied to optimize the data mining problems or to handle data directly. In population-based algorithms, individuals move through a solution space and search for solution(s) for the data mining task. The algorithm could be utilized to optimize the data mining prob-lem, e.g., the parameter tuning. The swarm intelligence algorithm could be directly applied to the data samples, e.g., subset data extraction. With the swarm intelligence, more effective methods can be designed and utilized in the massive data analytics problem.

In population-based algorithms, every solution is spread in the search space. Each solu-tion is also a data point; the distribution of solutions can be utilized to reveal the landscape of a problem. Data analysis techniques have been exploited to design new swarm intel-ligence/evolutionary algorithms, such as brain storm optimization algorithm [4, 5] and estimation of distribution algorithms [6]. In this paper, the population-based algorithms indicate the evolutionary computation algorithms and swarm intelligence algorithms. There are several existing solutions at the same time, and massive information is gener-ated over iterations. Thus, the big data analytics could be utilized to analyze the process of optimization. For non-population based techniques, such as neural networks, a large number of parameters are tuned in different layers, which may also be analyzed by data analytics techniques.

There are three key challenges in big data problems, the data modeling, computing model, and implementation platform. The mainstream of big data research includes computing model, such as deep learning [7], MapReduce [8], and Platform, such as Hadoop [9], and Apache Spark. The aim of this paper is to provide a comprehen-sive review of the optimization of population-based algorithms on data science and on the analysis of data science method for population-based algorithms. The remain-ing of the paper is organized as follows. “Data science” Section reviews the basic concepts of data science methods. “Population-based algorithms” Section reviews the general concepts of swarm intelligence and evolutionary algorithms and in particu-lar, four algorithms, which are particle swarm optimization (PSO), ant colony opti-mization (ACO), brain storm optimization (BSO), and fireworks algorithm (FWA). “Data science with swarm intelligence and evolutionary algorithms” Section reviews the swarm intelligence and evolutionary algorithms utilized to optimize data science methods and data analysis methods utilized to analyze swarm intelligence algorithms are reviewed.

II. RELATEDWORK

2.1 Data science

Currently, data science or data analytics is a popular topic in computer science and statistics. It concerns with a wide variety of data processing tasks, such as data collection, data management, data analysis, data visualization, and real-world applications. (clusters). The goal of clustering is to classify objects being similar (or related) to one another in the same cluster and put objects being distant from each other in different clusters [14].

Clustering is the process of grouping similar objects together. From the perspective of machine learning, the clustering analysis is sometimes termed as unsupervised learning. There are N points in the given input, D = {xi}Ni=1, the "interesting and/or useful pattern" can be obtained through similarity calculation among points [15].

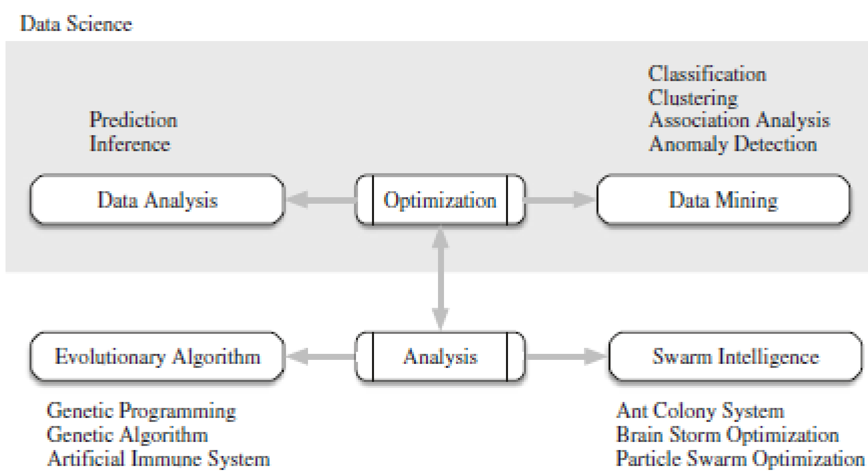

Fig. 1 Relationship between data science with evolutionary algorithm and swarm intelligence. Figure 1 gives a simple illustration on the connection between data analysis and swarm intelligence/evolutionary algorithms. Data science concerns the extraction of useful information from raw and massive data. It contains several processes on the data, such as collection, management, data analysis, model building, and visualization. Swarm intelligence is a relatively new subfield of computational intelligence which studies the collective intelligence of a group of simple individuals. In the swarm intelligence, the good optimization results could be obtained from the competition and cooperation of individuals. The aim of data science with evolutionary algorithms/swarm intelligence is to combine the strengths of evolutionary algorithms/swarm intelligence and data science techniques. It has two meanings: potential applications of evolutionary computation algorithms in data science and data analytics techniques in enhancing evolutionary computation algorithms. With the data analytics during the optimization process, the relationship between the algorithm and problems could be revealed. Evolutionary algorithms/swarm intelligence could be utilized to solve real-world big data analytics problems

2.2 Data Science with Statistics

Statistics is defined as a set of tools such as probability theory, real analysis, measure theory, asymptotic theory, decision theory, Markov chains, martingales, ergotic theory, etc., [16]. In contrast with data mining, the subject of statistics emphasizes much more on mathematical validation. In the field of statistics, the goal of data analysis is to explain data with relevant models, whereas the data mining or machine learning research is more interested in prediction with accuracy and thus does not concern much on the model's interpretability.

In this big data era, there is a big demand to improve traditional statistical data anal-ysis tools, which were used to study observations of instances of particular phenomena. A big data research paradigm claims for analyzing multi-source heterogeneous data in a theoretically sound and consistent way. To face challenges arising from big data analyt-ics problem, we need to integrate advanced techniques in disciplines like metaheuristics based optimization into the statistical toolbox.

As metaheuristics based optimization advocates naturally inspired strategy, whereas statistics stresses mathematical validation, a fusion of them can bring in a lot of new opportunities in proposing new models or algorithms. Some existing trials in fusion of statistics and optimization include the Bayesian simulation scheme mixed with computa-tional intelligence [17], the sequential Monte Carlo optimization methods [18], sequential Monte Carlo simulated annealing [19], sequential Monte Carlo samplers [20], population model-based optimization with sequential Monte Carlo [21], estimation of distribution algorithms [6], etc.

III. PROPOSED SYSTEM

We argue that one of the most promising directions in combining meta-heuristics and statistics is to handle the curse of dimensionality in high-dimension data analysis or optimization problems.

3.1 Intelligent transportation system

The traffic problem is arising in many cities now. Many factors affect the traffic and trans-portation system, e.g., the number of vehicles, weather, accident occurrence, etc., and the traffic information changes in real time. The purpose of intelligent transportation is to build more rapid, safe, and more efficient traffic and transportation system by con-structing the intelligent vehicles and road environment [22, 23]. In the intelligent trans-portation system, there are multiple objectives that ought to be satisfied. For instance, in the case of rapid transportation, environmental pollution, transportation schedul-ing: there are multi-objective conflicting objectives arising to achieve the predefined goal.

The intelligent transportation problem can be modeled as a large-scale, dynamical, and multi-objective optimization problem. The traditional methods have difficulties to solve this problem. Swarm intelligence has been proven to be an effective method in solving these problems [24] as no gradient information is necessary to find a feasible solution.

3.2 Population-based algorithms

Many real-world applications can be represented as optimization problems where algo-rithms are required to have the capability to search for the optimum. Most traditional methods can only be applied to continuous and differentiable functions [5]. The meta-heuristic algorithms are proposed to solve the problem, which the traditional methods cannot solve or, at least, be difficult to solve. Recently, kind of meta-heuristic algorithms, termed as swarm intelligence (SI) and evolutionary algorithm (EA), are attracting more and more attentions from researchers.

Swarm intelligence and evolutionary algorithms are collections of population-based searching techniques [1]. To search a problem domain, an SI or EA algorithm processes a population of individuals. Each individual represents a potential solution of the prob-lem being optimized. In EA or SI, an algorithm maintains and successively improves a collection of potential solutions until some stopping condition is met. The solutions are initialized randomly in the search space, and are guided toward the better and better areas through the interaction among solutions.

Swarm intelligence and evolutionary algorithms are two kinds of population-based and nature-inspired algorithm for optimization techniques. The difference is that swarm intel-ligence is roughly based on mimic of species interaction such as fish school, birds flock, and ant swarm, while evolutionary algorithms is roughly based on mechanisms of evolu-tion such as biological genetics and natural selection. The term "iteration" is used in SI and the "generation" is adopted in EAs to represent individuals at a time. The SI and EAs have the common paradigms in the optimization as follows.

- A population of individuals (potential solutions) is utilized in the search process.

- The "fitness" information, i.e., the "quality" of solutions, instead of function derivatives or other related knowledge are directly used in the search.

- The probabilistic, rather than deterministic, update rules are used to improve the "quality" of solutions.

As a general principle, the expected fitness of a solution returned should improve as the search method is given more computational resources in time and/or space. The data clustering analysis is a technique that divides data into several groups

More desir-able, in any single run, the quality of the solution returned by the method over iterations

should improve monotonically – that is, the fitness of the solution at time t + 1 should be no worse than the fitness at time t, i.e., fitness(t + 1) ≤ fitness(t) for minimization problems. The general procedure of swarm intelligence/evolutionary algorithms is given in Algorithm 1.

Algorithm 1: General procedure of swarm intelligence/evolutionary algorithms

1 Generate random solutions for an optimized problem, repair solutions if solutions violate any of the constraints;
2 Initialize all individuals in the population;
3 Evaluate all initialized individuals;

4 while the stopping criteria is not satisfied do
5 for all individuals in the population do
6 Reproduce individuals to form a new population;
7 Evaluate the fitness of each solution;
8 Select solutions with better fitness values;
9 Update non-dominate solutions in the archive;

There exist many swarm intelligence algorithms; among them most common ones are the particle swarm optimization (PSO) algorithm [30], which was originally designed for solving continuous optimization problems, and the ant colony optimization (ACO) algo-rithm, which was originally designed for discrete optimization problems [31]. The Brain Storm Optimization (BSO) algorithm and Fireworks algorithms (FWA) are two recently proposed swarm intelligence algorithms that are based on the convergence and diver-gence of solutions. Both BSO and FWA algorithms have the same operators: divergence operator and convergence operator; but two operators are utilized in different sequence.

3.3 Particle swarm optimization

Particle Swarm Optimization (PSO), which is one of the swarm intelligence techniques, was invented by Eberhart and Kennedy in 1995 [30, 32]. It is a population-based stochas-tic algorithm modeled on the social behaviors observed in flocking birds. Each particle, representing a solution, flies through the search space with a velocity that is dynamically adjusted according to its own and its companion's historical behaviors. The particles tend to fly toward better and better search areas over the course of the search process [1, 33].

In the particle swarm optimization problem, a particle not only learns from its own experience, but also learns from its companions. It indicates that a particle's 'moving posi-tion' is determined by its own experience and its neighbors' experience [34]. The general procedure of PSO algorithm is given in Algorithm 2.

3.4 Brain storm optimization

The brain storm optimization (BSO) algorithm was proposed in 2011 [4, 5], which is a young and promising algorithm in swarm intelligence. It is based on the collective behav-ior of human being, that is, the brainstorming process [4, 5]. The solutions in BSO are converging into several clusters. The best solution of the population will be kept if the newly generated solution at the same index is not better. New individuals can be gener-ated based on the mutation of one or two individuals in clusters. The exploitation ability is enhanced when the new individual is close to the best solution so far. While the explo-ration ability is enhanced when the new individual is randomly generated, or generated by individuals in two clusters. The procedure of BSO algorithm is given in Algorithm 4.

3.5 Data science with swarm intelligence and evolutionary algorithms

Data science concerns the extraction of useful information from raw and massive data. It contains several processes on the data, such as collection, management, data analy-sis, model building, and visualization. Swarm intelligence is a relatively new subfield of computational intelligence which studies the collective intelligence in a group of simple individuals. In the swarm intelligence, the good optimization results could be obtained from the competition and cooperation of individuals. Figure 1 gives a simple illustration on the connection between data analysis and swarm intelligence/evolutionary algorithms.

Algorithm 2: Procedure of fireworks algorithm

1 Initialize n locations;
2 while have not found "good enough" solution or not reached the pre-determined
maximum number of iterations do
3 Set off fireworks at n locations;
4 Generate sparks through explosion operator;
5Generate sparks through mutation operator;
6Obtain the locations of sparks;
7 Mapping the locations into feasible search space;
8Evaluate the quality of the locations;
9Select n locations as new fireworks;
Population-based algorithms in data science

Many real-world problems could be modeled as optimization problems, which have to find the optimum for one or several objective(s). The opportunities and challenges of evolutionary algorithms solving complex engineering optimization (CEO) problems are introduced in [39]. The optimization problems also happen in data mining tasks ortechniques, such as parameter tuning and minimum subset extraction. Swarm intelli-gence, especially particle swarm optimization or ant colony optimization algorithms, is utilized in data mining to solve single objective [40] and multi-objective problems [41]. Based on the two characters of particle swarm, the self-cognitive and social learning, the particle swarm has been utilized in data clustering techniques [42], document cluster-ing, variable weighting in clustering high-dimensional data [43], semi-supervised learning based text categorization, and the Web data mining [44]. The potential and possibilities of swarm intelligence algorithms in solving big data analytics problem were summarized in [45, 46]. Table 1 gives a list of applications that population-based algorithms solved problems in data science.

Classification Spam detection could be seen as a special case of classification prob-lems. Based on clonal principle in the natural immune system, the clonal particle swarm optimization (CPSO) algorithm was utilized in solving spam detection problems [49]. By cloning the best individual of successive generations, the CPSO algorithm could enlarge the area near the promising candidate solution and accelerate the evolution of solutions[49].A PSO algorithm based semi-supervised learning method was engaged to categorize Chinese text in [13]. A prototype generation method based on multiobjective PSO was developed to improve the performance of nearest neighbor classification techniques [50].

For other data mining problems, a PSO algorithm aided orthogonal forward regression was utilized in solving unified data modeling problem [51]; and a binary PSO algorithm proved effective to select the small subset of informative genes from gene expression data [52].

3.5 Artificial immune system

An artificial immune system uses principles in the operation of the human immune sys-tem and applies them to computationally intelligent systems. Artificial immune system (AIS) has been exploited to solve various data mining problems, especially classification or anomaly detection problems.

Classification A problem-oriented approach by designing an AIS algorithm for data mining, especially for classification, was advocated in [57]. An artificial immune system based multiclass classifier, named artificial immune system with local feature selection (AISLFS), was introduced in [58]. The local feature selection mechanism was embedded to reduce the dimensionality of the problem.

Anomaly detection Artificial immune system has been utilized to solve real-world anomaly detection problems, which are related to information security. A hybrid sys-tem based on artificial immune system and the self-organizing map was introduced to solve network intrusion detection problems [59]. An artificial immune system based virus detection system (VDS) was introduced in [60]. An artificial immune system based method for spam filtering was introduced in [61].

3.6 Genetic algorithms

Genetic algorithms have been utilized to solve many kinds of data mining problems, such as classification, clustering, Web mining, and association rules mining

Classification Genetic algorithms have been used to construct a compact fuzzy clas-sification system consisting of a small number of linguistic classification rules [62]. Two objectives are optimized by this fuzzy classification system, one is to maximize the number of correctly classified training patterns and the other is to minimize the number of selected rules. A genetic algorithm has been utilized to solve multi-label classification [63]. The goal of the multi-label classification task is to learn a classi-fier that predicts multiple class labels to an unlabeled instance based on features of an instance. A real-coded genetic algorithm (RCGA) was proposed to improve the classification performance of a polynomial neural network (PNN) [64]. The mean clas-sification accuracy (CA) is used as the fitness value of each solution for the training dataset.

Clustering Genetic algorithm has been used to optimize the clusters created during unsupervised clustering [65]. The solutions are consisted by hard partitions of the feature space, and the fitness function is a version of the hard c-means optimization function. In [66], genetic algorithms were utilized to search for the optimal, in the least squares sense, hierarchical clustering of a dataset. A multiobjective genetic algorithm based approach was proposed for fuzzy clustering of categorical data [67]. The fuzzy compactness and fuzzy separation of the clusters are optimized at the same time. A multiobjective algo-rithm, named DYNMOGA (DYNamic MultiObjective Genetic Algorithms) was proposed to solve community discovery problems in dynamic networks [68]. The aims of this pro-posed algorithm are to maximize cluster accuracy with respect to incoming data of the current time step, and to minimize clustering drift from one time step to the successive one. Two objectives are optimized by DYNMOGA, the first is the maximization of cluster accuracy, i.e., to maximize the snapshot quality of the current time step, and the second is the minimization of clustering drift, i.e., to minimize the temporal cost, which measures the distance between two clusters from one time step to the successive one. A genetic algorithm with spectral-based methodologies, named GANY was proposed to deal with the large data analysis problem [69]. The goal of GANY is generating a method to analyze more data using less resource.

Web mining Genetic algorithms were used for data-driven Web question answering problems, which need to find exact answers to natural languages (NL) questions. Answers are extracted directly from the N-best snippets, which have been identified by a standard Web search engine using NL questions [70].

Association rules mining Association rules mining problem is usually modeled as a multiobjective optimization problem. A literature reviews on multiobjective genetic algorithms and multiobjective genetic programming for rule knowledge discovery in data mining are given in [71]. An automated clustering method based on multiobjective genetic algorithms was proposed to solve fuzzy association rules mining problems [72]. The method is applied to decide on the number of fuzzy sets and for the autonomous mining of both fuzzy sets and fuzzy association rules. The goal of the method is to obtain a large number of item-set in less time by automatically cluster values of a quantitative attribute. A multiobjective genetic algorithm based approaches were utilized for mining optimized fuzzy association rules [73]. Two different forms of criterion are used: the one tries to determine the appropriate fuzzy sets of quantitative attributes in a pre-specified rule, which is also called as certain rule, and the other deals with finding both uncertain rules and their appropriate fuzzy sets.

3.7 Genetic programming

The genetic programming (GP) has the similar operators to the genetic algorithm (GA), which includes crossover, mutation and selection. The difference between genetic programming and genetic algorithm is that the GP has a population of tree-shaped indi-viduals, while the GA has a population of string-shaped individuals [1]. GP has been utilized to solve many kinds of data analysis problems.

Regression Symbolic regression based on Pareto Front GP is utilized to generate empir-ical models for industrial applications [74]. From the results of a small-sized industrial data set, the optimal settings of three parameters: the number of cascades, the number of generations, and the population size are tuned based on Pareto front GP.

Classification Performance bias may occur due to the unbalanced data sets, which may lead classifiers have good accuracy on the majority class, but very poor accuracy on the minority class. A multi-objective genetic programming (MOGP) is proposed, with accurate and diversified classifiers, performed satisfactory on both minority and majority classes [75, 76]. New fitness functions in GP for binary classification with unbalanced data are introduced in [77].

More literature about the evolutionary algorithms for clustering or data mining prob-lems could be found in [78–80].

3.8 Data analysis in population-based algorithms

The data mining techniques could be applied to design or analyze swarm intelligence algorithms. Massive information exists during the search process. For swarm intelli-gence/evolutionary algorithms, there are several individuals existing at the same time, and each individual has a corresponding fitness value. New individuals are generated as the iteration increases. There is also a massive volume of information on the "origin" of an individual, such as that an individual was created by applying which strategy and param-eters to which former individual(s). The data-driven evolutionary computation/swarm intelligence is a new approach to analyze and guide the search in evolutionary algo-rithms/swarm intelligence. These strategies could be divided into off-line methods and online methods. An off-line method is based on the analysis of previous storage search history, such as history based topological speciation for multimodal optimization [81] or maintaining and processing submodels (MAPS) based estimation of distribution algo-rithm on multimodal problems [82]. In comparison, for an online method, the parameters could be adaptively changed during the different search states.

The data modeling methods could be applied to inspire new swarm intelligence algo-rithms. In the brain storm optimization algorithm [5], every solution is spread in the search space. The distribution of solutions can be utilized to reveal the landscape of a problem. From the clustering analysis of solutions, the search results can be obtained. In the estimation of distribution algorithms [6], the space of potential solutions is explored by building and sampling explicit probabilistic models of promising candidate solutions.

Data clustering method could be utilized to improve the performance of swarm intel-ligence/evolutionary algorithms. A cluster and gradient-based artificial immune system is proposed to apply in optimization scenarios [83], and a clustering-based adaptive crossover and mutation probabilities for genetic algorithms are proposed in [84]. The number of clusters was analyzed in BSO algorithm solving different kinds of problems [85]. Based on clusters analysis, the search status could be obtained.

3.9 Freight prediction and recommendation system

International trade plays an important role for a country's economy. The parts/goods are transported among countries every day. A simple description of the business model for freight transportation by ships is shown in Fig. 2. The parts and goods are transported from a factory to another country through the freight forwarders and the carrier com-pany. From a cost perspective, shipping companies pay a price to the carriers or freight forwarders, then the carriers or freight forwarders add a markup for profit and might hire other forwarders. It should be noticed that there may be more than one freight forwarders for one goods.

Conclusions

In swarm intelligence and evolutionary algorithms, a population of individuals is utilized to evolve the optimized functions or goals by cooperative and competitive interaction among individuals. Massive information exists during the search process, such as the dis-tribution of individuals and the fitness of each solution. To improve the search efficiency or to recognize the search state, the data generated in the optimization process should be analyzed.This paper has reviewed the connection between data science and swarm intelli-gence/evolutionary algorithms. The potential combination of data science and swarm intelligence/evolutionary algorithm in optimization and data analytics was also analyzed. Data science involves prediction or inference on a large amount of data. Swarm intelli-gence studies the collective behaviors in a group of individuals. With the combination of data science, swarm intelligence and evolutionary algorithms, more rapid and effective methods can be designed to solve optimization and data analytics problem.

REFERENCES

1. Kennedy J, Eberhart R, Shi Y. Swarm Intelligence. San Francisco: Morgan Kaufmann Publisher; 2001.
2. Dorigo M, Stützle T. Ant Colony Optimization. Cambridge: MIT Press; 2004.
3. Eberhart R, Shi Y. Computational Intelligence: Concepts to Implementations. San Francisco: Morgan Kaufmann Publisher; 2007.
4. Shi Y. Brain storm optimization algorithm In: Tan Y, Shi Y, Chai Y, Wang G, editors. Advances in Swarm Intelligence.

Lecture Notes in Computer Science, vol. 6728. Berlin Heidelberg: Springer; 2011. p. 303–9.

5. Shi Y. An optimization algorithm based on brainstorming process. Int J Swarm Intell Res (IJSIR). 2011;2(4):35–62.
6. Pelikan M, Goldberg DE, Lobo FG. A survey of optimization by building and using probabilistic models. Comput Optim Appl. 2002;21(1):5–20.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. Nautre. 2016;521:436–44.

8. Dean J, Ghemawat S. Mapreduce: Simplified data processing on large clusters. In: Proceedings of 6th Symposium on Operating Systems Design and Implementation (OSDI 2004); 2004. p. 137–49.
9. White T. Hadoop: The Definitive Guide 4th edn. Sebastopol: O'Reilly Media, Inc; 2015.
10. Donoho DL. 50 years of data science. Technical report, Stanford University. 2015.
11. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Mag. 1996;17(3): 37–54.
12. Cervantes A, Galván IM, Isasi P. AMPSO: A New Particle Swarm Method for Nearest Neighborhood Classification. IEEE Trans Syst Man Cybern B Cybern. 2009;39(5):1082–91.
13. Cheng S, Shi Y, Qin Q. Particle swarm optimization based semi-supervised learning on Chinese text categorization. In: Proceedings of 2012 IEEE Congress on Evolutionary Computation (CEC 2012). Brisbane, Australia: IEEE; 2012.
p. 3131–198.
14. Tan PN, Steinbach M, Kumar V. Introduction to Data Mining. Boston: Addison Wesley; 2005.
15. Murphy KP. Machine Learning: A Probabilistic Perspective. Adaptive computation and machine learning series.Cambridge: The MIT Press; 2012.
16. Friedman JH. Data mining and statistics: What's the connection? In: Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics; 1997. p. 1–7.
17. Liu B, Ji C. A general algorithm scheme mixing computational intelligence with Bayesian simulation. In: Proceedings of the 2013 Sixth International Conference on Advanced Computational Intelligence; 2013. p. 1–6.
18. Liu B. Posterior exploration based sequential Monte Carlo for global optimization. Technical report, Nanjing University of Posts and Telecommunications. 2015.
19. Zhou E, Chen X. Sequential monte carlo simulated annealing. J Glob Optim. 2013;55(1):101–24.
20. Del Moral P, Doucet A, Jasra A. Sequential monte carlo samplers. J R Stat Soc Ser B Stat Methodol. 2006;68(3):411–36.
21. Chen X, Zhou E. Population model-based optimization with sequential monte carlo. In: Proceedings of the 2013Winter Simulation Conference: Simulation: Making Decisions in a Complex World. Washington: IEEE; 2013.p. 1004–15.
22. Kohata N, Sato M, Yamaguchi T, Baba T, Hashimoto H. Evolutionary computation for intelligent agents based on chaotic retrieval and soft DNA In: McKay B, Yao X, Newton CS, Kim J-H, Furuhashi T, editors. Simulated Evolution and Learning. Lecture Notes in Computer Science, vol. 1585. Berlin Heidelberg: Springer; 1999. p. 251–9.
23. Teodorovic´ D. Transport modeling by multi-agent systems: A swarm intelligence approach. Transp Plan Technol. 2003;26(4):289–312

24. Li X, Yao X. Cooperatively coevolving particle swarms for large scale optimization. IEEE Trans Evol Comput. 2012;16(2):210–24.
25. Chui M, Löffler M, Roberts R. The internet of things. McKinsey Q. 2010;2:1–9.
26. Atzori L, Iera A, Morabito G. The internet of things: A survey. Comput Netw. 2010;54(15):2787–805.
27. Liu Y, Zhou G, Zhao J, Dai G, Li XY, Gu M, Ma H, Mo L, He Y, Wang J, Li M, Liu K, Dong W, Xi W. Long-term large-scale sensing in the forest: recent advances and future directions of greenorbs. Front Comput Sci China. 2010;4(3):334–8.

28. Kulkarni RV, Venayagamoorthy GK. Particle swarm optimization in wireless-sensor networks: A brief survey. IEEE Trans Syst Man Cybern Part C Appl Rev. 2011;41(2):262–7.

29. Kulkarni RV, Förster A, Venayagamoorthy GK. Computational intelligence in wireless sensor networks: A survey. IEEE Commun Surv Tutor. 2011;13(1):68–96.

30. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks (ICNN 1995); 1995. p. 1942–1948.

31. Dorigo M, Maniezzo V, Colorni A. Ant system: optimization by a colony of cooperating agents. IEEE IEEE Trans Syst Man Cybern B Cybern. 1996;26(1):29–41.

32. Eberhart R, Kennedy J. A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science; 1995. p. 39–43.

33. Eberhart R, Shi Y. Particle swarm optimization: Developments, applications and resources. In: Proceedings of the 2001 Congress on Evolutionary Computation (CEC2001); 2001. p. 81–6.

34. Cheng S, Shi Y, Qin Q. Population diversity of particle swarm optimizer solving single and multi-objective problems. Int J Swarm Intell Res (IJSIR). 2012;3(4):23–60.

35. Tan Y, Zhu Y. Fireworks algorithm for optimization In: Tan Y, Shi Y, Tan KC, editors. Advances in Swarm Intelligence.Lecture Notes in Computer Science, vol. 6145. Berlin Heidelberg: Springer; 2010. p. 355–64.

36. Tan Y. Fireworks Algorithm: A Novel Swarm Intelligence Optimization Method. Berlin Heidelberg: Springer; 2015.

37. Cheng S, Qin Q, Chen J, Shi Y, Zhang Q. Analytics on fireworks algorithm solving problems with shifts in the decision space and objective space. Int J Swarm Intell Res (IJSIR). 2015;6(2):52–86.

38. Martens D, Baesens B, Fawcett T. Editorial survey: swarm intelligence for data mining. Mach Learn. 2011;82(1):1–42.

39. Chai T, Jin Y, Sendhoff B. Evolutionary complex engineering optimization: Opportunities and challenges. IEEE Comput Intell Mag. 2013;8(3):12–15.