Structure learning of bayesian network using swarm intelligent algorithm: a review

Shahab Wahhab Kareem [a,b,1,*], Shavan Askar [a,2], Kosrat Dlshad Ahmed [a,3]

[a] Information System Engineering Department, Technical Engineering College, Erbil Polytechnic University, Erbil, Iraq
[b] College of Engineering and Computer Science, Lebanese French University, Kurdistan, Iraq
[1] shahab.kareem@epu.edu.iq*; [2] shavan.askar@epu.edu.iq; [3] kosrat.ahmed@epu.edu.iq
* corresponding author

ARTICLE INFO

Article history
Received August 2, 2021
Revised August 20, 2021
Accepted August 30, 2021

Keywords
Bayesian Network
Conditional Independence Test
Structure Learning
Global Search
Local Search
Pigeon Inspired Optimization

ABSTRACT

Machines using Bayesian networks can be used to construct the framework of information in artificial intelligence that connects the variables in a probabilistic way. "Deleting, reversing, moving, and inserting" is an approach to finding the best answer to the proposition of problem in the algorithm. In the Enhanced Surface Water Searching Technique, mostly, the hunt for water is done by elephants during dry seasons, It is Pigeon Optimization, Simulated Annealing, Greedy search, and the BDeu metrics being reviewed in combination to evaluate all these strategies being used in order to solve this problem. They subjected different data sets to the uncertainty matrix in an investigation to find out which of these approaches performed best. According to evaluation data, the algorithm shows stronger results and delivers better points. Additionally, this article also represents the structure learning processes for Bayesian Network as well.

1. Introduction

Biologists and social scientists build relational structures for their study [1]. But the network layout of a lot of nodes is very dense. Sometimes used in applied in statistics and machine learning, the probabilistic model is a random variable between an independent framework and a classification model, probabilistic networks are more often used to produce structure. Different models express the likelihood distributions in multidimensional space in different forms [1]. A distinction in terms of structure and composition is the form of probability distribution is used to differentiate between the two. A Bayesian model is more than just a reflection of information, it is a particular type of probability graph. When there are connections between different random variables, it will show interdependence [2]. Although an acyclic structure is described as random network between parameters, it can also be thought of as a guided acyclic graph.

A Bayesian network has more often represents a multidimensional distribution or configuration of curves rather than arrows since it uses directional edges to express the probability distribution." High-dimensional modelling will also cost a great deal of money to do. One approach to construct network topology is by score and process is another route dependent on restriction. It is difficult to create score functions, as well as difficult to check for the optimization space. The primary advantage of the methodology based on restrictions is how to accurately determine whether a kind of freedom has been achieved. When learning is concerned with network optimization, the best search approach is to use a scoring system that allows matching of the highest degree of structure on results [3]. Based on the results of the variables being independent of each other, the network can find a way to create a Bayesian framework. So far, several clever Bayesian network layout learning algorithms have been suggested. That has well-clusters of observations (a frequently used Bayesian network built around LFO observations) (PC). Practical, feasible have various advantages and drawbacks, and thus, this algorithm has the network of being used in a wide range of applications, such as in a mixed learning situation. "Compensate for network search space constraints with learning algorithm and recover Bayesian network score from sparse graph" (CIBNS). Using the Bayesian network to help, and data that we were given as an example above, we can cut down on the blindness of the quest [4].

As a result, a combination of constraint-action and search-based methods will further increase the performance of Bayesian network research. Around the same time, it's good for finding the global optimum. Since the relationships of random variables can be represented with direct acyclic graphs (DAGs) and therefore useful in realms such as artificial intelligence, medicine, bioinformatics, and sociology, it has been widely applied in the economy, as well. The framework learning mechanism has gained popularity in recent years and is now almost ubiquitous [5]. With regards to learning the structures of DAGs from evidence, two methods have predominated: As well, the quest and restriction and mixed methods have been used. Rebuilding DAGs Conditional approaches for determining BN structure, such as the use of independence measures as well as search-based methods have been suggested [6]. Constraint-based methods construct using variables that are conditional independence (correlation) dependent on each other He uses IC, PC, and TPDA (three-oriented dependency analysis) in this class. A score-based search is used to discover the network's ideal setup under which search constraints are applied only to the super structure of the graph. A reduction in the search space will result in better solutions [4]. Verma and Pearl gave an IC which looks for all subsets of u and v, for the assumption that no subsets of the two variables are conditionally dependent on some other subset of variables and all other subsets are independent on that subset of the whole, and which partition S into u and v are mutually exclusive and which are mutually nonexclusive. This procedure was suggested by Spirtes and Glymour, an iterative way of looking for the cardinality of increasing subsets of items was [3]. The PC algorithm constrains the set of vertices that may separate U and V to vertices that are either directly connected or that are adjacent to V [7]. As the number of vertices in a DAG expands, the asymptotic consistency of the algorithm remains intact.

2. Related Works

2.1. Bayesian Network

Because of their capacity to deal with nondeterministic variables in the physical universe, probabilistic models have been widely used in recent years. The joint probability distribution of random variables can be defined by a Bayesian network (Bayesian) network (BN) [8]. A Bayesian network is a graph in which variables are represented by nodes and Bayesian probabilities are described. Structure learning is based learning is just a form of parameter learning. Likelihood is obtained for each variable probabilistically for those variables it has to have [9]. It detects a DAG with one node for each variable in the structure. Variables used in the system are conditionally independent of their non-descendants with regard to their values, whether the conditions specified are not true for any of the nodes lower in the network [8]. lately, extensive research has been conducted on discovering how BNs work by actual evidence. Learned dependency graph analysis may be a valuable in helping to define the problem which is also utilized in solving it (Friedman et al. 1999). The Cooper-Herskovitzler k2 algorithm was first suggested in 1992, and virtual annealing BN was improved in 1995 (SA). Genetic algorithm (GA) is used to apply it to the design of the neural network structures.

A Bayesian network S = {G, θ} for a series of n quantities that expresses the conditional probability mass function of its constituent variables and θ is defined by a guided acyclic graph X = {X1, X 2,..., X n} etc.

$$p(G|D) \propto p(G)p(D|G) \quad (1)$$

In the Bayesian network, G, a correlation between two variables implies that they are interconnected, and an edge occurs. As if the structure G is valid, then it means that prior to seeing some data, the prior data holds true. It is conditional on the probability. To treat P (G and D) as a ranking [10], one can use a heuristic search algorithm, one can search for a high-scoring network. For e.g., a greedy search the area in question, looking at all nearby structures, calculates the score, and goes to the structure with the best score. It ends because the new arrangement is greater than any of any of the surrounding structures. Hence, computing a network structure can be described as an optimization problem [11], where the objective is to find the best possible quality for the training data. It can be made using a Bayesian method, minimal knowledge criterion, or a combination of the two.

These measurements may be applied to entire networks, and their results summarized as a (or composed) as the total (or product) of their parts. It's easier to score and therefore to locate when it's done locally. If you are looking for familiarity, K2, HillClimber, SA, and GA are good ways to go. Probabilistic models, because of their power to describe uncertain information, are often highly effective aids. Probability theory offers an analysis method for determining how the elements are linked, to ensure the system's consistency. There is supposed to be little overlap between the two sets of findings and different approaches are necessary to come up with new models for the data. Graph theory provides an intuitive general-purpose interface for creating arrays of interactional [12], variable-structure data that can be used in varied algorithms. Maybe 'probabilistic graphical simulations' may be referred to as Bayesian networks.

While probabilistic networks are a more complex and more flexible technique, Bayesian networks (BN) can be considered a simpler analytical approach for machine learning [4]. They can be applied through information development, discourse, and derivation. Bayesian network had the structure: DAGs with two critical nodes. A variety of the network is created by the parameters and its overall configuration. This structure describes parameter dependencies, and these parameters represent conditional probabilities [9]. Without an effective search tool, building a Bayesian network is a challenge. Given a dataset, the optimization problem of finding the optimum configuration of a Bayesian network (BN) is NP-complete [4]. Although significant research to come up with an approximation of network structures has been done In essence, there are two approaches to structural learning in Bayesian networks. While the first uses constraint-based techniques [13], and searches, the second employs a method that utilizes scoring and searching techniques. A calculation is used to find the BN composition.

2.2. Structure Learning of Bayesian Network

A probabilistic paradigm incorporates mathematics, as well as graphical representations of concepts that were developed by graph theory [14]. There are two problems that must be addressed: ambiguity and unpredictability. Graphical models are joint probabilistic systems that depict random dependency relationships Algorithms provide a large impact in analyzing and in the design and development of machine learning approaches Diagrammatic models use probabilities, where nodes represent random variables and/association between pairs of variables may be drawn to display relationships. compressed mutual likelihood distributions There are two essential kinds of software programs: writer-prompts and reader-prompts. Markov Random Fields (MRFs) or Bayesian Networks may also be classified as undirected models. Observational learning has been researched throughout the latest decade. The use of experience or observed evidence can be used to generate new principles [15]. Although there has been some study on this topic, different approaches have been tried Make just the changes to the data necessary to reconstruct the feature. To change the whole graph to optimize a ranking, but the alterations could create dependencies. The aim of optimization techniques is to create a local version of the network to better approximate the global structure. Most analysis with Bayesian networks is limited to scenarios in which the configuration does not change. Traditional algorithms also used scores were used for the bulk of these algorithms. We may determine the likelihood of all DAG models if there are just a few random variables, including all possibilities [12].

$$f(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i) n > 2 \tag{2}$$

For all practical purposes, this amount should be huge. Checkering has shown that for certain groups of DAG-patterns that this is really is so. Using smarter search algorithms may help solve this sort of issue. For example, we will select values of a stochastic variable (SV) that increase the probability of obtaining p (should be higher than one maximizing model.) As the number of variables grows, finding the optimal models among all possible models becomes infeasible to the method.

2.3. Pigeon Inspired Optimization

Currently, the majority of Pigeon-inspired optimization (PIO) has been used in continuous optimization problems. as the follow-up from problem orientation: the Metropolis adoption of simulated annealing is suggested as a solution to this discrete PI-O (DPIO) algorithm; (TSPs).

$$Vi(t) = Vi(t-1). e^{-Rt} + rand. (Pg - Pi(t-1)) \tag{3}$$

$$Pi(t) = Pi(t-1)Vi(t) \tag{4}$$

DPIO is developing a new chart and compass that will make it easier for explorers to understand and advance. a modern hallmark operator, which is intended to increase the performance of TSPs using cooperative learning and heuristic knowledge. The Metropolis approval criteria is used to determine if new solutions should be allowed to converge or not, particularly if they are otherwise trying to be effective [16]. Experiments were carried out to observe the behavior of the map and compass operator as well as other landmark workers and learn their individual differences.

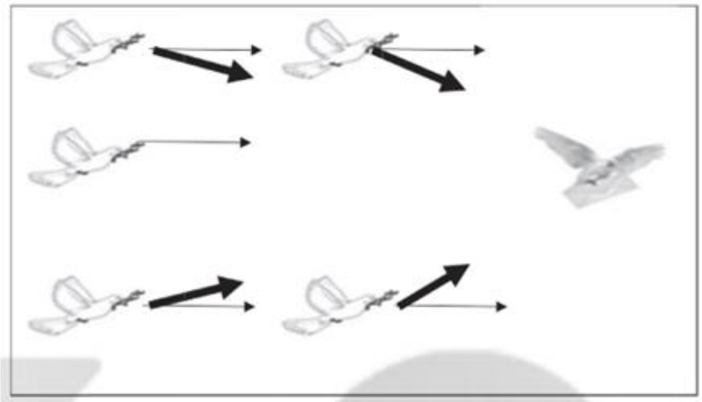

Fig. 1. Compass operator and map for PIO.

Pigeon Inspired Optimization is a method is a novel bio-inspired optimization method In terms of pigeons [17], it was on the first account that the navigational methodology was created, and in air-vehicle applications it was later.

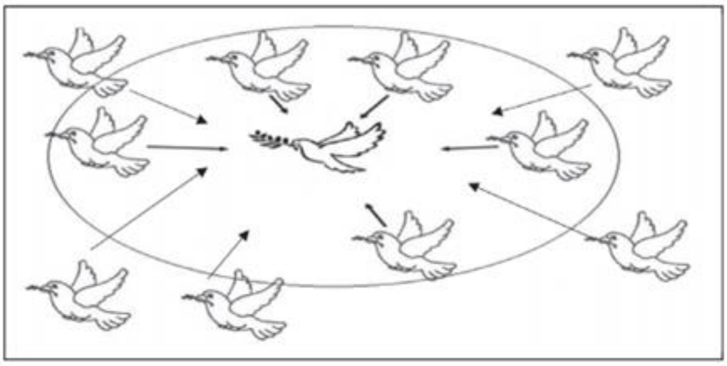

Fig. 2. Landmark operator model for PIO.

Any PIO has two essential-valued components: a mapping driver and a way finder, and a geographic/direction driver, and another that values landmarks. In the map and compass domain driver, there are many suns, while in the landmark driver, there are just a few landmarks [16]. In layered form, it mimics the aptitude of handyman do vessel pigeons during the map and compass, each team member learns new navigation techniques, including D-dimensional explorations. We will get to know the new position and velocity IP and velocity I by each one of our techniques. Much of the pigeons are able to locate the targets because of the magnetic field. They often use landmarks to figure out where they are.

The suggested solution utilizes a Bayesian search using a PIO search tool [18]. To calculate the Bayesian network configuration, the BDeu metric was used. Any pigeon in the population encodes a location and velocity in the specified room [19]. This place is believed to be the optimal search location. It is focused on two methods. For the first technique, the local quest procedure is navigated using a schematic of a globe and compass. Second, it employs a landmark paradigm for world-wide quest [18]. Pseudo code of this algorithm is seen in Fig. 3. PIO algorithm may include the use of various neighborhood-specific search algorithms. When searching for solutions created by pigeons, you can depend on being up-to-to-date. The DAGs for studying BRTs are a part of the solution space to be investigated. Any pigeon represents a solution and is started on an empty seed DAG [20]. When the pigeon is looking for a likely network layout, defined as the BDeu metric, it explores inside the quest room. Equation 4 seeks to find the network with the greatest BDeu score.

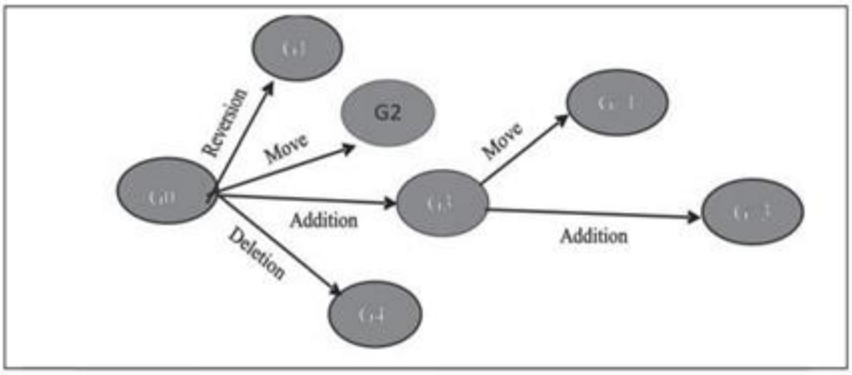

Fig. 3. Compass steps for a particular pigeon.

At any turn, new ideas are developed through trial and error. If you start with an empty graph, all arcs will be appended. The process is repeated until the specified number of arcs has been created. As a result, each operator is initialized with the initial population and the one that raises BDeu score is chosen [21]. As long as BDeu or as long as necessary, the operator sees an improvement, the pigeon can keep flying [22]. Operators usually are called the four basic optimization operators: addition, elimination, movement, and mutation, and replacement. It is an uncomplicated setup in this particular domain and only replaces one competitor edge per time, with the four operators working around the board. On the other side, moving the solution about in some medium affects it, in a mild manner [23]. Correcting a solution without the three basic operators would not automatically result in a better solution [21]. Higher ground-hopping occurs with proximity to a sought-after destination. Moving the cursor to a lot about is referred to as pigeon-skipping in literature. Landmark is a complex and interesting combination of travelling and seeking. A DAG defines a pigeon as one which does revert, moves, makes new arcs, and ads, and reaches solutions G1, G2, G3, and G4 in succession [24]. As suggested by G3, then considering the BDeu score, it will look for another operator that satisfies G3 If the BDeu (behavioral + economics) score of G+1 is greater than that of G, than it can carry out the associated action.

3. Method

Using positivist philosophical approach, this research targets to maintain the development of knowledge to assess the outcomes. In this process, gradual management of the information and process simulation is necessary. For the comparative analysis of the literature, several determiners such as AND, OR, BUT etc. have been used for the keywords such as Bayesian network, structure processing, structure learning, conditional independence test etc.

To simulate the comparative review, articles have been collected from the reliable journals. Articles published from 2010 to 2020 have been chosen for the comparative review because of the scarcity of quality journals. Systematic comparison among the selected articles will be conducted in this review. Comparative analysis have been conducted based on available articles in this regard.

4. Result and Analysis

Through developing theories of ambiguity reasoning in the 1980s, Bayesian networks (or Bayesian belief networks) have come to the forefront of artificial intelligence science. recently, BNs have shown to be an important method in dealing with unpredictable systems and data processing of multivariate random variables Determining the network configuration and parameters with datasets derived from previous analyses. Structure or parameter learning into two parts: structure learning and parameter identification. In the former, the network topology determination involves identifying the topologically suitable sample sets. The details of the network topology are needed to establish the network parameters. Bayesian learning also involves access to network topology and data, making it the foundation of Bayesian network growth. A good way to find the 'optimization structure' is to be based on identifying effective structural principles.

There are two models of Bayesian Networks, which could be derived from a Bayesian framework learning process, as a way of breaking down a Bayesian inference system, the DAG method may be considered a graph that codifies the interdependence between variables [7]. conditional freedom is confirmed by experimentation. Stress on relevancy approaches Once you have chosen a quest strategy, the best network will be found [7]. There are means of assessing how accurately we can classify participants (commoner ones include maximum probability, Bayesian Information Criterion, [31] Bayesian ranking, and Minimum Description Length), as well as uncommon means (Maximum Product-Likelihood, Bayesian Information, and Minimum Description). Searching for the most optimal network is an intractable challenge and refined algorithms are almost impossible to implement [7]. Manually controlled approaches have included techniques, such as Greedy hill-climbing [31], Simulated Annealing, Evolutionary Algorithm (EA), and Genetic Algorithm (GA). GA and EA have been widely employed. Because the number of variables is high, both of the above algorithms are likely to return a local optimised network structure until convergence is reached [32].

Table.1 Findings from related previous research

Source	Findings
[25]	In unpredictable situations, Bayesian Networks (BNs) may be good methods for creating models of ideas and reasoning. Bayesian Network configuration is deemed a difficult challenge to learn from a dataset because of the search space difficulty. This paper introduces a novel approach for structure learning that is focused on PSO (Particle Swarm Optimization) and the K2 algorithm. PSO is used to check the space of orderings in order to better understand a bayesian network's layout. Each network's fitness is determined by running the K2 algorithm and comparing the result to the fitness of the network. Our solution gives greater efficiency than the other approaches.
[26]	The paper proposes a hybrid algorithm that they term a Bayesian network model-based conditional independence test and a heuristic scan (CIBNS, CI-based Bayesian Network Search). To ensure the solution consistency and search speed, the latest algorithm initially uses the conditional independence test to compact the search space. After that, the algorithm implements a heuristic search method that incorporates the BDeu Measure score in order to boost performance in the construction process. On simulated and actual info, the latest algorithm finds that it is capable of efficiently constructing a network for aviation. It outclasses hill climbing and local hunt in terms of utility and precision.
[27]	Extending the representational capacity of continuous Bayesian networks beyond exponentially-distributed state transformations has long been done through the use of phase-type distributions. A framework for learning phase-type distributions was introduced in this article. To efficiently obtain good phase-type approximations for a variety of parametric distributions, we use particle swarm optimization to reduce a changed KL divergence value.
[28]	The authors discussed a NP-hard problem for Bayesian network architectures from results in this paper. A proposal using shared knowledge and PC algorithm techniques was introduced in this article. This algorithm uses shared knowledge to find the original undirected graph. You can firstly use the PC algorithm to find a PDAG. Based on our experiments, it seems that our algorithm is more effective than the PC algorithms under the same conditions.
[29]	The authors regarded Bayesian networks structure learning is widely considered to be a difficult computational problem, as the more candidate graphs you have, the more the ratio of variables grows exponentially. There have been several heuristic searching techniques that attempt to improve network layout. We demonstrate two methods for enhancing Bayesian network layout learning using search optimization in this article. This method combines elements of Greedy quest with elements of Bee optimization algorithms. To test this suggested strategy, we are going to use two search techniques.
[30]	Using Bayesian networks, machine learning experts may construct a framework for information that depicts the probability relationship between variables. Elephant Swarm Water Search Algorithm (ESWSA) was introduced as a new way to build Bayesian networks. The algorithms involved include the ones mentioned below: Deleting, Reversing, Inserting, and Moving are both used in order to arrive at the optimum solution structure with the ESWSA. Mostly, elephants' use of water quest technique during droughts is incorporated into the ESWSA algorithm. To assess all algorithms, BDeu (Bird-inspired Efficient) score function is used. They compared the confusion matrix performances of these methods with different data sets to get a better understanding of them. The findings of the evaluations show that the algorithm suggested does better than the other algorithms, both in terms of efficiency and results.

In this paper, a novel approach focused on a new-to-discrete Particle-Behavioural Quantum algorithm (NDPB) is used to learn BNs' structure Statistical dependencies are easy to express in a Bayesian network (BN). It harnesses the power of graph theory to represent random variables' Dynamic Bayesian Networks (DBNs) is a type of Bayesian Networks (BNs) which deals with time-varying processes. Due to DBN's ability to describe nonlinear, time-dependent, changing, and probabilistic relationships, much of DBN learning and modelling research has emerged. DBNs also been employed in many different sectors. Many complex Bayesian networks are built by asking a professional for help. When working for smaller networks, elicitation may be a straightforward, but when doing so for huge numbers of variables, it becomes cumbersome and time-consuming. If information is accessible, it is feasible to construct the model on that [32].

If all the data have been collected, studying two networks can be modelled when the complex Bayesian network results into the expression of prior network and transformation network since the fitness mechanism is a closed form [32]. Data normally include errors, because we typically cannot accurately observe the mechanism we are trying to represent. An additional difficulty with missing data is that it operates as a closed form with comprehensive data, though not limited to that. What's needed for learning a database structure from incomplete data is entirely is much more effort than what's required for a well-structured database (He et al. 2018). from 1998 onward, the use of DBN (Structural EM), can be seen in latent variable models. Deterministic approaches, on the other hand, are vulnerable to finding local optima, are said to be weak at generalizing, may be wide in their quest space [33]. When a local maximum has been reached, an easy solution to the problem is to use a stochastic approach this paper introduces complex Bayesian networks using particle swarm optimization (PSO) [34]. We choose that as a result of the job. Since networks can be viewed as components, we can develop more complex systems by sharing their component parts with higher fitness.

5. Conclusion

Using the Pigeon-Inspired Learning method, we concentrated on the structure learning issue and applied it to the Bayesian network. They are useful for the design of information representation systems in machine learning. It is possible to encode probabilistic dependencies among the variables using Bayesian networks. Scoring and scan is one layout learning methodology at Bay Networks. Propose a new Bayesian network framework based on PIO (PIO). It's a basic concentration rate. An unbelievable and remarkable navigational skill can be seen in pigeons in the way they just seem to find their way a guided acyclic graph Any chart has a fitness score that tells its ability to demonstrate this reality.

The algorithm takes time to explore the solution space, so it is performed using a landmark, compass, and map operator before it achieves the best or a suitable structure is found. When implementing the suggested approach, simulated annealing and greedy search were contrasted with each other with the BDeu value. In addition, we tested the uncertainty approaches in a variety of data sets. A specific algorithm yields outstanding results as shown by the effects produced it produces better results, is as effective, and produces higher and better values than simulated annealing and greedy algorithms. We used the score and search method, where PI as a parameter and the search tool. a search strategy that uses the random flights of a pigeon as its basis.

The hunt seems to follow the path of least resistance, so PIO is a popular technique for finding discrete solutions. Easily tailored to a specific to every target area. Prompting PIO exploration results in a likely-to-to-be-be-tried solutions because of allowing pigeons to travel in short range leads to plausible solutions by parameter input. According to the suggested approach, the structure of the network, it can find outstanding structure solutions, which means it can measure better function and approximates more accurately. The algorithms make the quest faster and convergence is rapid. The suggested solution may be studied as a study in parallel processing.

References

[1] X. Sun, C. Chen, L. Wang, H. Kang, Y. Shen, and Q. Chen, "Hybrid Optimization Algorithm for Bayesian Network Structure Learning," *Information*, vol. 10, no. 10, p. 294, Sep. 2019, doi: 10.3390/info10100294.

[2] J. Xin, G. Chen, and Y. Hai, "A Particle Swarm Optimizer with Multi-stage Linearly-Decreasing Inertia Weight," in *2009 International Joint Conference on Computational Sciences and Optimization*, Apr. 2009, pp. 505–508, doi: 10.1109/CSO.2009.420.

[3] C. Yuan, B. Malone, and X. Wu, "Learning optimal Bayesian networks using A* search," 2011.

[4] P. O. Djan-Sampson and F. Sahin, "Structural learning of Bayesian networks from complete data using the scatter search documents," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 2004, pp. 3619–3624, doi: 10.1109/ICSMC.2004.1400904.

[5] C. Yang, J. Ji, J. Liu, J. Liu, and B. Yin, “Structural learning of Bayesian networks by bacterial foraging optimization,” *Int. J. Approx. Reason.*, vol. 69, pp. 147–167, Feb. 2016, doi: 10.1016/j.ijar.2015.11.003.

[6] X.-S. Yang, “A New Metaheuristic Bat-Inspired Algorithm,” in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, 2010, pp. 65–74.

[7] Y. Zhang, J. Liu, and Y. Liu, “Bayesian Network Structure Learning: The Two-Step Clustering-Based Algorithm,” 2018.

[8] J. Cowie, L. Oteniya, and R. Coles, “Particle swarm optimisation for learning bayesian networks,” in *ICCIIS 2007, World Congress on Engineering, WCE 2007*, 2007, pp. 71–76.

[9] Shao-Zhong Zhang and Lu Liu, “MCMC samples selecting for online bayesian network structure learning,” in *2008 International Conference on Machine Learning and Cybernetics*, Jul. 2008, pp. 1762–1767, doi: 10.1109/ICMLC.2008.4620690.

[10] J. Wang and S. Liu, “Novel binary encoding water cycle algorithm for solving Bayesian network structures learning problem,” *Knowledge-Based Syst.*, vol. 150, pp. 95–110, Jun. 2018, doi: 10.1016/j.knosys.2018.03.007.

[11] G.-G. Wang, S. Deb, and L. dos S. Coelho, “Elephant Herding Optimization,” in *2015 3rd International Symposium on Computational and Business Intelligence (ISCBI)*, Dec. 2015, pp. 1–5, doi: 10.1109/ISCBI.2015.8.

[12] J. Li and J. Chen, “A Hybrid Optimization Algorithm for Bayesian Network Structure Learning Based on Database,” *J. Comput.*, vol. 9, no. 12, pp. 2787–2791, 2014.

[13] L. Antonio Pereira Silva, J. Batista Nunes Bezerra, M. Barbosa Perkusich, K. Costa Gorgônio, H. Oliveira de Almeida, and A. Perkusich, “Continuous Learning of the Structure of Bayesian Networks: A Mapping Study,” in *Bayesian Networks - Advances and Novel Applications*, IntechOpen, 2019.

[14] T. Rahier, S. Marié, S. Girard, and F. Forbes, “Fast Bayesian Network Structure Learning using Quasi-Determinism Screening,” in *JFRB 2018-9èmes Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes*, 2018, pp. 14–24.

[15] S. Tong and D. Koller, “Active learning for structure in Bayesian networks,” in *International joint conference on artificial intelligence*, 2001, pp. 863–869.

[16] A. J. Kulkarni, G. Krishnasamy, and A. Abraham, *Cohort Intelligence: A Socio-inspired Optimization Method*, vol. 114. Cham: Springer International Publishing, 2017.

[17] Y. Cao, M. Raoof, E. Szabo, J. Ottosson, and I. Näslund, “Using Bayesian Networks to Predict Long-Term Health-Related Quality of Life and Comorbidity after Bariatric Surgery: A Study Based on the Scandinavian Obesity Surgery Registry,” *J. Clin. Med.*, vol. 9, no. 6, p. 1895, Jun. 2020, doi: 10.3390/jcm9061895.

[18] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, “Learning Bayesian networks from data: An information-theory based approach,” *Artif. Intell.*, vol. 137, no. 1–2, pp. 43–90, May 2002, doi: 10.1016/S0004-3702(02)00191-1.

[19] T. Koski and J. Noble, *Bayesian networks: an introduction*. John Wiley & Sons, 2011.

[20] A. R. Khanteymoori, M.-H. Olyaee, O. Abbaszadeh, and M. Valian, “A novel method for Bayesian networks structure learning based on Breeding Swarm algorithm,” *Soft Comput.*, vol. 22, no. 9, pp. 3049–3060, May 2018, doi: 10.1007/s00500-017-2557-z.

[21] M. Bartlett and J. Cussens, “Integer Linear Programming for the Bayesian network structure learning problem,” *Artif. Intell.*, vol. 244, pp. 258–271, Mar. 2017, doi: 10.1016/j.artint.2015.03.003.

[22] K. M. Salama and A. A. Freitas, “ABC-Miner: An Ant-Based Bayesian Classification Algorithm,” in *Swarm Intelligence*, 2012, pp. 13–24.

[23] F. Liu, F. Tian, and Q. Zhu, “Bayesian Network Structure Ensemble Learning,” in *Advanced Data Mining and Applications*, 2007, pp. 454–465.

[24] S. Sencer, E. Oztemel, O. Torkul, C. Kubat, H. Taskin, and G. Yildiz, “Bayesian Structural Learning with Minimum Spanning Tree Algorithm,” 2013.

[25] S. Aouay, S. Jamoussi, and Y. Ben Ayed, “Particle swarm optimization based method for Bayesian Network structure learning,” in *2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*, Apr. 2013, pp. 1–6, doi: 10.1109/ICMSAO.2013.6552569.

[26] Z. Long, “A Bayesian Network Based Structure Learning Algorithm,” in *2016 International Conference on Robots & Intelligent System (ICRIS)*, Aug. 2016, pp. 12–15, doi: 10.1109/ICRIS.2016.2.

[27] L. J. Perreault, M. Thornton, R. Goodman, and J. W. Sheppard, “A Swarm-Based Approach to Learning Phase-Type Distributions for Continuous Time Bayesian Networks,” in *2015 IEEE Symposium Series on Computational Intelligence*, Dec. 2015, pp. 1860–1867, doi: 10.1109/SSCI.2015.259.

[28] J. Li and Y. Zhang, “A Method for Learning Bayesian Network Structure,” in *2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics*, Aug. 2014, pp. 222–225, doi: 10.1109/IHMSC.2014.156.

[29] S. Kareem and M. C. Okur, “Bayesian Network Structure Learning Using Hybrid Bee Optimization and Greedy Search,” in *3rd International Mediterranean Science and Engineering Congress*, 2018, pp. 1–7.

[30] S. Cheng and Y. Shi, *Handbook of Research on Advancements of Swarm Intelligence Algorithms for Solving Real-World Problems*. IGI Global, 2020.

[31] B. Andrews, J. Ramsey, and G. F. Cooper, “Scoring Bayesian networks of mixed variables,” *Int. J. Data Sci. Anal.*, vol. 6, no. 1, pp. 3–18, Aug. 2018, doi: 10.1007/s41060-017-0085-7.

[32] S. B. Gillispie and M. D. Perlman, “Enumerating Markov Equivalence Classes of Acyclic Digraph Models,” *arXiv Prepr. arXiv1301.2272*, Jan. 2013, [Online]. Available: http://arxiv.org/abs/1301.2272.

[33] C.-C. He and X.-G. Gao, “Structure Learning of Bayesian Networks Based On the LARS-MMPC Ordering Search Method,” in *2018 37th Chinese Control Conference (CCC)*, Jul. 2018, pp. 9000–9006, doi: 10.23919/ChiCC.2018.8483049.

[34] A. H. Gandomi, X. S. Yang, S. Talatahari, and A. H. Alavi, *Metaheuristic applications in structures and infrastructures*. Newnes, 2013.