Middle-East Journal of Scientific Research 18 (9): 1343-1348, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.18.9.12375

Structure Learning of Bayesian Belief Networks Using Simulated Annealing Algorithm

Alireza Sadeghi Hesar

Department of Computer Engineering,
Mashhad Branch, Islamic Azad University, Mashhad, Iran

Abstract: Basically, Bayesian Belief Networks (BBNs) as probabilistic tools provide suitable facilities for modelling process under uncertainty. A BBN applies a Directed Acyclic Graph (DAG) for encoding relations between all variables in state of problem. Finding the beststructure (structure learning) ofthe DAG is a classic NP-Hard problem in BBNs. In recent years, several algorithms are proposed for this task such as Hill Climbing, Greedy Thick Thinning and K2 search. In this paper, we introduced Simulated Annealing algorithm with complete details as new method for BBNs structure learning. Finally, proposed algorithm compared with other structure learning algorithms based on classification accuracy and construction time on valuable databases. Experimental results of research show that the simulated annealing algorithmis the bestalgorithmfrom the point ofconstructiontime but needs to more attention for classification process.

Key words: Bayesian Belief Network · Simulated Annealing Algorithm · Structure Learning Algorithms

INTRODUCTION

In recent decade, building expert systems for modrithmelling under uncertainty has been one of the main problems in the field of data mining sciences. Different tools for dealing with uncertainty have been developed such as Dempster-Shafer theory and fuzzy framework. Today, Bayesian Belief Networks (BBNs) are very popular data mining techniques to encode relationships among a set of variables in conditions of uncertainty. SinceBBNsarebased onMathematical logic and probabilities theory, in comparisonwiththe fuzzy logic have strongerreasoning tools. Briefly, a BBN is graphical model that presents Joint Probability Distribution (JPD) of dependentvariables in the state of problem. Finding the best structure for a BBN is a NP-Hard challenge. So,use ofsearchalgorithms with approximate answersis an inevitable task. So far, manymodelshave been proposedinthe field of structure learning of BBNs using heuristic and metaheuristic algorithms.One of theclassicstudiesinthis area is paper of R. Etxeberria *et al*. (1997). For the first time, they evaluated behaviour of genetic algorithms for structure learning of Bayesian networks from finite dataset. The results show that genetic algorithms are not appropriate tools for learning of BBNs when problem domain is very large. Yoichi Motomura and Isao Hara (2004) introduced a BBNstructure learning method using Artificial Neural Networks (ANNs). After training considered ANN, it can encode the conditional probabilities between all variables. Ioannis Tsamardinos *et al*. (2006) presented a new method to find structure called Max-Min Hill Climbing or MMHC. This method is based on three different approaches from search-and-score techniques, local learning and constraint-based learning. MMHC constructs skeleton of BBN and then applies score-based greedy hill-climbing algorithm to orient the edges. Jun-Zhong JI *et al*. (2009) presented an improved method based on the conditional probabilities and ant colony optimization to find the BBNs topology. The results of research on the benchmark sets described that the proposed method is very efficient in large scale databases. A. R. Khanteymoori *et al*. (2009) [1] apply Tabu search algorithm for this task. Tabu search is an iterative search algorithm that uses a local search method at each cycle to find for the best station. They in 2011 proposed new approach for BBNs structure learning based on asexual reproduction optimization (ARO). ARO in this paper was introduced as an evolutionary-based method that mathematically describes the budding mechanism of asexual reproduction. In final, ARO was

Corresponding Author: Alireza Sadeghi Hesar, Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

applied to model two real problems. Results show ARO algorithm is very faster than genetic algorithm because ARO has faster operator [2]. In this paper, we will introduce, a new structural learning method using Simulated Annealing or SA algorithm. SA algorithm is a local search method, means we search the next state from the set of neighborhood states and then decide to move to that state or not [3]. The rest of the paper is organized as follows [4]. The Bayesian Belief Networks and key structure learning methods will be introduced in the section 2 [5]. Section 3 describesthe proposedmethod with full details. Finally, Section 4 summarizes the main results and conclusions [6].

Bayesian Belief Networks: Bayesian belief networks or BBN is named based on studies of Thomas Bayes (1702-1761) in the field of probability theory. His studies led to the production of Bayes' rule which is expressed as follows [7-15]:

☻ (1)

The BBN asaprobabilisticstructure factorizes the Joint Probability Distribution (JPD) of a set of random variables using Observationaldata [16]. The BBNs apply a Directed Acyclic Graph (DAG) to encode relationships between a set of variables in state of problem. The nodes of DAG represent discrete variables and arcs show Conditionalindependenceof variables. Each node of DAG has a Conditional Probability Table (CPT) that presents probability of each state of node according to any combination of parent states. The JPD computation in Bayesian Networks can be expressed mathematically using Equation 2:

$$\mathrm{P_r}\left(Y_1, Y_2, \ldots, Y_n\right) \prod_{i=1}^{n} P\left(\frac{Y_i}{\pi_i}\right) \tag{2}$$

Equation 2 shows that the joint probability distribution for node Y in DAGis product of the probability of each Y_i of node Y given the parents of Y_i.

Consider the following simple example that indicates some of the properties of BBNs by DAG and with CPTs for each node when the variables are discrete (consists of eight random variables). This network is designed for a diagnosis problem.

Structure Learning: The BBNs construction process can be separated two major steps: parameter learning and structure learning. Since the BBNs are statistical model

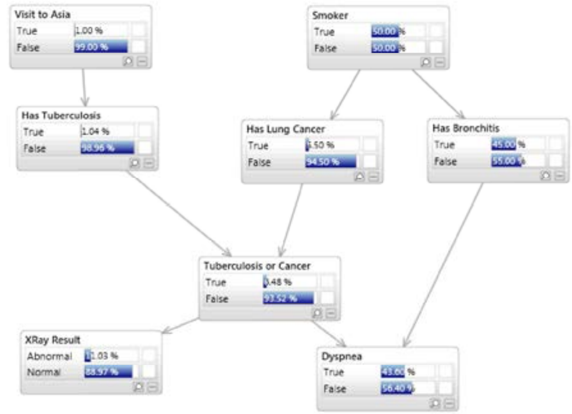

Fig. 1: A Simple BBN for Diagnosis Problem

and the parameter learning is a method for learning in statistics, it can also be used in Bayesian networks. Basically, parameter learning is calculation of the conditional probabilities for the obtained structure of BNN and parameters in a BBN are the probabilistic values in the CPTs. The most effective methods for parameter learning are Reduce Gradient (RG) and Maximize Likelihood Estimation (MLE).The main objective of the structure learning is finding the best structure for BBN that is compatible with existing dataset and is optimal from the point of complexity and construction tine. The structure learning includes two different approaches of constraint-based learning and score-based learning. In constraint-based learning the network structure is achieved using conditional independence relations between variables. But score-based learning assigns a score to each possible topology and tries to maximize it using metric scoring functions. The finding optimal structure for BBNs is a NP-Hard problem. To solve this problem, Greedy search algorithms such as K2 search, hill-climbing and Tabu search are a common choice. Generally, greedy search algorithms are based on score-based methods that offer scale or metric solutions. These methods evaluate all of the possible relationships between nodes in the general space and determine an instance with the maximum ranking.

K2 Search Algorithm: K2 algorithm is a score-based algorithm that finds the best possible structure using an iterative process between all possible topologies. The scoring function of K2 algorithm for i nodes is in this format:

$$\left(\mathrm{i}, \pi_i\right) = \coprod_{j=1}^{|\Phi_i|} \frac{\left(r_i - 1\right)}{\left(Ni_j + r_i - 1\right)} \prod_{k=1}^{r_i} \alpha_{ijk} \tag{3}$$

I is the current node, r_i is the number of states, π_i is the parent of, $|\varphi_i|$ is the number of values within the CPT of, α_{ijk} is The number of cases in the dataset in which has its k^{th} value and have their j^{th}value in CPT and N^{ij} is sum of α_{ijk} for each state of i. Before the execution of scoring function, the variables must to be ordered and a fixed and limited amount to be considered for the parents of each node. Usually in specific applications, the experts of field determine order of nodes and amount of parents for each node. This process prevents from generating loops in the graph and final score of network will obtains by multiplying the individual score of nodes.

Hill Climbing Algorithm: The Hill Climbing algorithm is an optimization method based on local search and score-based methods. Basically,the problemsusing this methodarediscussed that have severalanswers and main challenge is selecting the best answer among all the results. Hill climbing attempts to maximize or minimize a target function, where considered element is a discrete or continuous variable. At each cycle of execution, algorithm will adjust a single element in ☻ and determine whether the change improves the ☻ value. This algorithmh as a fund amental objection and sticks in functions that have very local minimum or maximum points such as Ackley function (Figure 2).

Greedy Thick Thinning: Greedy Thick Thinning or GTT algorithm is a constraint-based learning method for BBNs. The GTT algorithm first produce a null graph based on mutual probabilities among variables. After that, GTT adds an edge if its connected nodes are not conditional independent. In final step, provided graph is reviewed again and the edges related to independent but connected nodes will be removed from graph. Briefly, the GTT algorithm is an existing topology optimizer using modifying the structure and dependence relations.

Simulated Annealing Algorithm: Simulated annealing algorithm is a probabilisticmetaheuristic method for the optimization problems in a large search space. Thisalgorithmhas been developedin 1983 by Scott Kirkpatrick and Daniel Gelatt. Basically,most of the metaheur is ticalgorithms have been formed based on bench marking and Simulation of the law sorrelation ships in the nature. Simulated annealing algorithmis also established based on annealing processof metals. The annealing process, in the first step increases metals temperature using extra heat and in next step imposes Cooling process and the gradual temperaturer eduction on

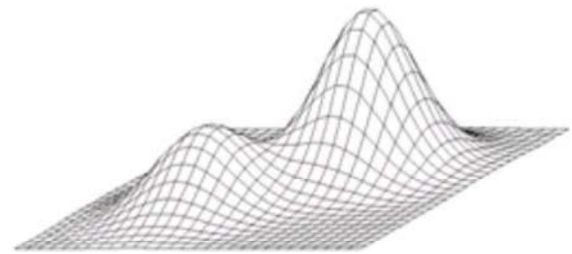

Fig. 2:

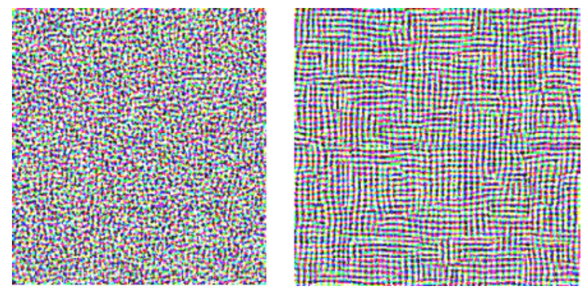

Fig. 3: Changing Patterns of Annealed Metal Atoms

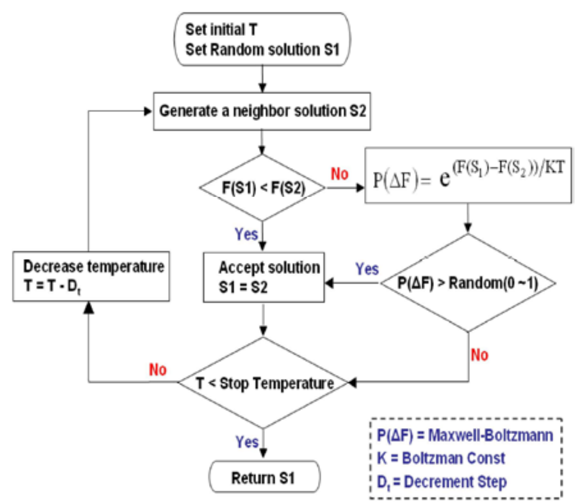

Fig. 4: Flowchart of Simulated Annealing Algorithm

them. In this process, fusionof metalincreasesthe speed of atoms movement greatly. Next, a gradualde crease of temperatur ecauses the formation of specific patterns in the adoption of it satoms (Figure 3). The changing patterns of annealed metalatoms make valuable properties such as strength of metals. Figure 4 presents flowchart of simulates annealing algorithm.

Asshownin Figure 5, simulated annealing is developed based on local search. Therefore, the designing the appropriate local search methods related tothe conditions and limitations of simulated problemsin thisalgor ithmisa very important issue. Generally, after analysing the algorithm, its advantages can beintroducedas follows:

```
s ← s0; e ← E(s)
sbest ← s; ebest ← e
k ← 0
while k < kmax and e > emax
  T ← temperature(k/kmax)
  snew ← neighbour(s)
  enew ← E(snew)
  if P(e, enew, T) > random() then
    s ← snew; e ← enew
  if e < ebest then
    sbest ← snew; ebest ← enew
  k ← k + 1
return sbest
```

Fig. 5: Pseudo code of Simulated Annealing Algorithm [2]

- Considered algorithm consumesvery little memory. (Unlike thegenetic algorithm that hasa high consumption)
- Its implementation is simple simple. (Compared with the methods of its class)
- This algorithm produces the acceptable answersin state of problem due tofocus on the local search.
- Because of guided random process, simulated annealing algorithm (the lower probabilityof accept ance fornon-optimal responses) has theability totransition froma local optimum.
- Simulated annealing as a guided method for optimization problems can produce better results in TSP[1] problem. (Compared with the genetic algorithm)
- Staticandlow-coststructure of algorithm is similarto Steady-State genetic algorithms or SSGA.
- But disadvantages of simulated annealing algorithm can be listed as follows:
- Theinterest method isvery dependent on the values of parameters.

If the wrongvalue be selected for the in itial temperature, the algorithmgets stuckin the localoptimum.

RESULTS AND DISCUSSIONS

In this section, we run three described algorithms and Simulated Annealing algorithm as proposed method on six different databases in different fields and properties. Table 1 presents these databases with complete details.

WBCD (Wisconsin Diagnostic Breast Cancer) database is results of the Fine Needle Aspiration (FNA) test that has been achieved in Medical and Research Center of Wisconsin (MRCW) by Dr. William H. Wolberg in 1992-1995. Considered classes of WBCD are Benign and Malignant (seriousness probability ofbreastcancer). Thesecond database shows diagnosing of cardiac Single Proton Emission Computed Tomography (SPECT) images. Each instance is classified into two categories: normal and abnormal. This data set was collected in Medical College of Ohio to extract main features from the original SPECT images. The Precipitation dataset (Database 3), contains 9000 instances from rate of dailyrainfall in Texas province (1985-2009). There are four different classes: dry (rf[3]<3), light-rain (3= rf <10), moderate-rain (11= rf <20) and heavy-rain (rf =20). (All measured in millimeter.) The Average of temperature data set (Database 4), contains 4500 instances. There are five classes for it: very cold (tm[4]<2), cold (2= tm <12), moderate (12= tm <25), warm (25= tm <40) and hot (tm =40). (All measured in centigrade). Shuttle Landing Control data set is a small database of all nominal values with two classes of Normalconditions and heat conditions that extracts rules for manual control of the spacecraft by auto landing. The Iris dataset (Final Database), is the most popular dataset applied in artificial intelligence, contains 150 examples each of three types of flower: Iris setosa, Iris versicolor and Iris virginica. There are four attributes for it: petal length, petal width, sepal length and sepal width (all measured in centimeter).Test and evaluation processes are executed a total of 5 times on a computer with a 2.6 GHz Core 2 CPU and 2 GB of RAM memory. We recorded classification accuracy and

Table 1: Experimental Data sets

No	Title	Field of Database	Number of Instances	Number of Attributes	Number of Class
1	WBCD	Physiology	559	32	2
2	SPECT Heart	Physiology	267	22	2
3	Precipitation	Meteorology	9000	12	4
4	Ave. Temperature	Meteorology	4500	12	5
5	Shuttle Control	Space Science	15	6	2
6	Iris	Botany	150	4	3

[1]Traveling Sale man Problem
[2]http://en.wikipedia.org/wiki/Simulated_annealing
[3]Rainfall
[4]Temperature

Table 2: Classification Accuracies of four Algorithms

	K2 Search	Hill Climbing	Greedy Tick Thinning	Simulated Annealing
Database 1	91.86	71.16	83.63	93.07
Database 2	92.24	83.25	88.75	96.42
Database 3	84.54	67.65	80.15	84.54
Database 4	87.34	71.92	71.48	86.13
Database 5	96.42	88.97	88.97	96.42
Database 6	89.14	80.19	89.14	89.14

Table 3: Construction Times of four Algorithms

	K2 Search	Hill Climbing	Greedy Tick Thinning	Simulated Annealing
Database 1	0.08	0.12	0.1	0.08
Database 2	0.07	0.11	0.1	0.1
Database 3	1.02	1.78	1.34	1.13
Database 4	0.11	0.65	0.72	0.54
Database 5	0.03	0.04	0.04	0.04
Database 6	0.08	0.13	0.12	0.12

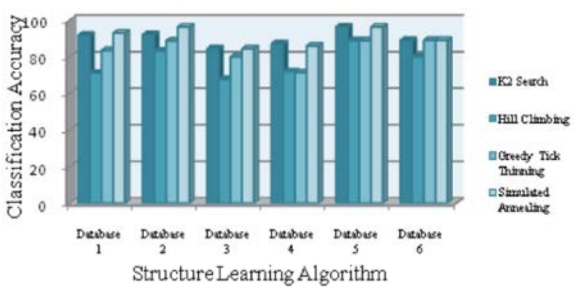

Fig. 6: Classification Accuracies of four Algorithms

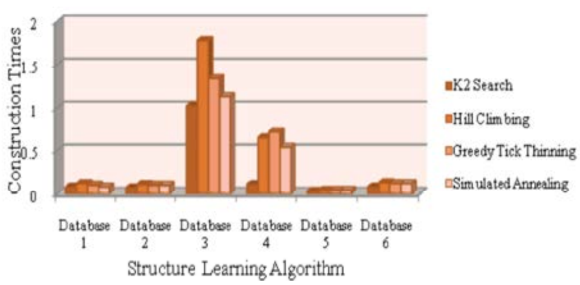

Fig. 7: Construction Times of four Algorithms

construction time for each algorithm separately. Figure 6 and Table 2 present accuracies of the learning algorithms discussed in this paper.

Table 3 and Figure 7 present construction times of the all learning algorithms discussed in this paper.

CONCLUSIONS

The Bayesian Belief Network as aprobabi listics tructure factorizes the Joint Probability Distribution (JPD) of a set of random variables using Observationaldata under uncertainty. The BBNs construction process can be separated two major steps: parameter learning and structure learning. Finding the best structure for a BBN is a NP-Hard problem. So far, several algorithms were proposed for this task such as Hill Climbing, Greedy Thick Thinning and K2 search. In this paper, we introduced Simulated Annealing algorithm as new score-based method for BBNs structure learning. Finally we compared proposed method with other structural learning methods on six different databases. As a result, it was shown that from the point of classification accuracy, the Simulated Annealing is the best algorithm and K2 is less time consuming compared to other algorithms.

REFERENCES

1. Khanteymoori, A., M.M. Homayounpour and M.B. Menhaj, 2009. A Bayesian Network Based Approach For Data Classification Using Structural Learning, Communications in Computer and Information Science, Springer, pp: 25-32.
2. Cofino, A., R. Cano, C. Sordo and J.M. Gutierrez, 2000. Bayesian Networks for probabilistic Weather Prediction, Proceedings of the 15th European Conference on Artificial Intelligence, IOS Press, pp: 695-700.
3. Cano, R., C. Sordo and J.M. Gutierrez, 2004. Applications of Bayesian Networks in Meteorology Advances in Bayesian Networks, in Advances in Bayesian Networks, Gámez et al. eds. Springer, pp: 309-327.
4. Kent, M., H. Le, M. Tadross and A. Potgieter, 2008. Weather Forecasting With Bayesian Network, Ph.d Thesis, Department of Computer Science University of Cape Town, Cape Town, South Africa.
5. Lee, B. and J. Jacob, 1998. Learning a Probabalistic Model of Rainfall Graphical Models, Quarterly Journal of the Royal Meteorological Society, pp: 90-96.
6. Korb, K. and A. Nicholson, 2004. Bayesian Artificial Intelligence, Chapman & Hall/CRC Press LLC, pp: 75-83.
7. Lamma, E., F. Riguzzi and S. Storari, 2005. Improving the K2 Algorithm Using Association Rule Parameters, Elsevier B.V, 1-11, Dipartimento di Ingegneria, Universit`a di Ferrara, Via Saragat 1, 41100 Ferrara, Italy.
8. Heckerman, D., 1999. A tutorial on learning with Bayesian networks, In M. I. Jordan (Ed.), Learning in Graphical Models, pp: 301-354.

9. Russell, S. and P. Norvig, 2003. Artificial intelligence, a modern approach, New York: Prentice Hall.
10. Friedman, N. and M. Goldszmidt, 1996. Learning Bayesian networks with local structure, Proc. The Twelfth Conference on Uncertainty in Artificial Intelligence, pp: 252-262.
11. Heckerman, D., D. Geiger and D.M. Chickering, 1995. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learining, 20(3): 197-243.
12. Sanchez, L., J. Casillas and O. Cordon, 2002. Some relationships between fuzzy and randomset-based classifiers and models, International Journal of Approximate Reasoning, Elsevier, pp: 175-203.
13. Ahmad, F.K., S. Deris and N.H. Othman, 2002. The inference of breast cancer metastasis through gene regulatory networks, Journal of Biomedical Informatics, 45(2): 350-362.
14. Maskery, S.M., H. Hu, J. Hooke, C.D. Shriver and M. N. Liebman, 2008. A Bayesian derived network of breast pathology co-occurrence, Journal of Biomedical Informatics, 41(2): 242-250.
15. Ramírez, N., H.G. Mesa and H. Calvet, 2007. Diagnosis of breast cancer using Bayesian networks Computers in Biology and Medicine, 37(11): 1553-1564.
16. Kurt Omurlu, I.. K. Ozdamar and M. Ture, 2009. Comparison of Bayesian survival analysis and Cox regression analysis in simulated and breast cancer data sets, Expert Systems with Applications, 36(8): 11341-11346.