

Probabilistic Graphical Models for the Medical Industry Developed Using Enhanced Learning Algorithms

Chih-Chiang Wei*

Department of Digit Fashion Design, Toko University, Taiwan

Abstract

This study presents two enhanced learning algorithms used for discovering probabilistic graphical models based on the Bayesian Network (BN) structure. The two heuristic structure learning algorithms, namely Tabu Search (TS) and Simulated Annealing (SA), were empirically evaluated and compared regarding efficiency. These algorithms were applied to real-life data sets for the vertebral column. A data set containing values for six biomechanical features was used to classify patients into three categories, namely, Disk Hernia (DH), Spondylolisthesis (SL), and Normal (NO), and two categories, namely, Abnormal (AB) and NO. The results indicated that SA is a more effective algorithm than TS. However, the empirical results obtained using TS indicated that the TS algorithm is promising because of its relatively simple network structure.

Keywords: Probabilistic graphical model; Classification; Medicine

Introduction

The earliest artificial intelligence approaches to medical diagnosis were based on Bayesian and decision-theoretic schemes. Difficulties in obtaining and representing quantities of numbers and both the computational and representational complexity of probabilistic schemes caused a long-lasting departure from these approaches [1]. In numerous domains, such as genetics, clinical diagnoses, direct marketing, finance, and online business, data sets arise with thousands of variables and a low ratio of cases to variables. Such data present dimensional difficulties for classifying a target variable [2] and identifying critical predictor variables [3]. Hence, determining influence (i.e., the causal relationships between the target variable and predictor variables) is a great challenge.

Recently, the development of probabilistic graphical models, such as Bayesian Networks (BNs) and closely related influence diagrams has renewed interest in applying probability theory in intelligent systems. BNs are successfully applied to various problems, including those involved in machine diagnosis, user interfaces, natural language interpretation, planning, vision, robotics, and data mining [1]. In medical diagnosis, the elimination of redundant tests can reduce the risks to patients and lower health care costs [4].

BNs are graphical knowledge representation tools. Every probability distribution can be defined using a BN [5]. As a result, BNs are widely used in problems in which uncertainty is managed using probabilities [6]. However, identifying the optimal network structure requires searching through all possible structures [7]. Although promising results have been obtained using global search techniques, their computation cost renders them unfeasible [8,9]. Heuristic methods, such as Simulated Annealing (SA) [10,11] and Tabu Search (TS) [12,13] can be used for large combinational optimization problems. This BN learning approach, using evolutionary techniques, can be employed to identify a high-scoring BN structure. This paper describes two enhanced learning algorithms (TS and SA) used to develop BN structures based on data. This study considered a medical case involving pathologies of the vertebral column.

Experiment

The vertebral column is a system composed of a group of vertebrae, invertebrate discs, nerves, muscles, medulla, and joints. This complex

system can exhibit dysfunctions that cause backaches with various intensities. The data set applied in this work was supplied by Dr. Henrique da Mota, who collected the data during a medical residence in spine surgery [14]. This data set contains data on 310 patients obtained from sagittal panoramic radiographies of the spine. Each patient is represented in the data set by six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine (Table 1). In this experiment, the patients in the first subdivision of the data set were classified into one of three categories: Disk Hernia (DH) (60 patients), Spondylolisthesis (SL) (150 patients), and Normal (NO) (100 patients). For the second subdivision of the data set, the DH and SL categories were merged into a single category labeled "Abnormal" (AB). Thus, the patients in the second subdivision were classified into one of two categories: AB (210 patients) and NO (100 patients).

BN and learning algorithm

Theory of BNs: The probabilistic graphical model is a classification method based on Bayes' theorem [15]. A Bayesian network $B = \langle G, \theta \rangle$ is a directed acyclic graph G that models probabilistic relationships among a set of random variables $U = \{X_1, \dots, X_n, \Omega\}$, where n denotes the number of attributes, X_n is the attribute term, and Ω is the

Notation	Description	Mean	Min-max
PI	Pelvic incidence	60.50	26.15~129.83
PT	Pelvic tilt	17.54	-6.55~49.43
LA	Lumbar lordosis angle	51.93	14.00~125.74
SS	Sacral slope	42.95	13.37~121.43
PR	Pelvic radius	117.92	70.08~163.07
GS	Grade of Spondylolisthesis	60.50	-11.06~418.54

Table 1: Six attributes in vertebral column dataset.

*Corresponding author: Chih-Chiang Wei, Department of Digit Fashion Design, Toko University, Taiwan, Tel: 886-953077710; E-mail: d89521007@ntu.edu.tw

Received October 28, 2013; Accepted November 23, 2013; Published November 28, 2013

Citation: Wei CC (2013) Probabilistic Graphical Models for the Medical Industry Developed Using Enhanced Learning Algorithms. Ind Eng Manage 2: 119. doi: 10.4172/2169-0316.1000119

Copyright: © 2013 Wei CC. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

class variable. Each variable in U has specific states or values denoted by lower-case letters $\{x_1, \dots, x_n, \omega\}$. The symbol Θ represents the set of parameters used to quantify the network. Each node contains a local probability distribution given its parents. Additional information is provided in Friedman et al. [16] and Pernkopf and O’Leary [17].

Learning algorithm: A basic condition for general search algorithms to be applicable to a combinatorial optimization problem (f, S) , where f is the cost function and S is a solution space, is that the cost function f can be easily calculated for elements of the neighborhood of a solution. An appropriate neighborhood structure that fulfills this condition should be chosen [18].

(a) Tabu search: TS are a general-purpose combinatorial optimization algorithm [19]. The basic idea is to begin with an arbitrary solution and recursively select a new solution from the neighborhood of the previous solution that maximally increases a predefined cost function. If there is no solution that increases the cost function, the present solution is a local optimum. The search process leaves this local optimum by selecting a solution that minimally reduces the cost function. When a solution is selected that reduces the cost function, immediate reselection of the local optimum visited previously is prevented by maintaining a list of solutions that are forbidden, which is called the “Tabu list.” The length (tll) of this Tabu list determines how much iteration are required to return to a local optimum; the higher the tll is, the more iterations are required to possibly return to the local optimum, and the higher the probability is that it is not visited again. The solution returned by TS is the optimal solution visited during the execution of the algorithm. TS is attractive for its simplicity, the few parameters that are necessary, and the empirical results, which are promising for application in a variety of combinatorial optimization problems [20, 21]. The tll and a stop criterion depend on the instance of the combinatorial optimization problem (f, S) . Details on the TS algorithm are provided in Bouckaert [18].

(b) Simulation annealing: SA is a general-purpose combinatorial optimization algorithm that was introduced by both Kirkpatrick et al. [22] and Cerny [23]. SA was motivated by the process of annealing metal to harden it. Annealing begins by heating metal until it melts and all particles move freely. The temperature is then reduced carefully to allow the particles to arrange themselves in a highly structured lattice [18]. In SA, a problem begins with an initial solution that is iteratively modified according to control parameter T , which is analogous to temperature [24]. As the parameter T is reduced, the convergence criterion becomes increasingly difficult to satisfy. Finally, if T is lowered sufficiently, no additional changes in the solution space are possible. To avoid being frozen at a local optimum, the SA moves slowly through the solution space. This controlled improvement of the objective value is accomplished by accepting non improving moves with a certain probability that decreases as the algorithm progresses. This criterion is a Boltzmann’s probability distribution (Metropolis criterion) as a function of temperature T :

$$p(\Delta F) = \exp\left(\frac{-\Delta F}{T}\right) \quad (1)$$

Where F is the objective function, ΔF is the increment of the objective function (i.e., $\Delta F = F(v') - F(v_i)$), v_i is current values, and v' is a randomly generated new solution in the neighborhood of v_i . A detailed description can be found in Kirkpatrick et al. [22], Kalivas et al. [25], and Swierenga [26].

Parameter training: The cross-validation sub sampling approach

was employed to evaluate the BNs. The entire data set was randomly partitioned into 10 equal-sized subsets. During each run, one of the partitions was chosen for testing, and the remaining partitions were used for training. This procedure was repeated 10 times so that each partition was used for validation exactly once, making it a 10-fold cross-validation.

(a) Tabu search: TS involve hill climbing until an optimum is reached. The initial network used for structure learning is a naive BN, a network with an arrow from the classifier node to every other node. The main parameter tll was chosen by conducting a sensitivity analysis in the interval [1]. The True-Positive Rate (TPR), False Positive Rate (FPR), and precision were calculated, and the testing results are plotted in (Figure 1). As shown in the figure, the suitable tll was 6.

(b) Simulated annealing: This BN learning algorithm uses SA to identify a high-scoring network structure. The factor by which the temperature decreases in each iteration was set to 0.999. The number of iterations to be performed by the SA search was set to 1,000. The start temperature of the SA search was the main parameter. The start temperature determines the probability that a step in the wrong direction in the search space is accepted. The TPR, FPR, and precision were calculated by conducting a sensitivity analysis of the start temperature in the interval [1], and the testing results are plotted in (Figure 2). As shown in the figure, the suitable start temperature was 11.

Model Evaluation

Definition of performance

To provide the most comprehensive description of diagnostic

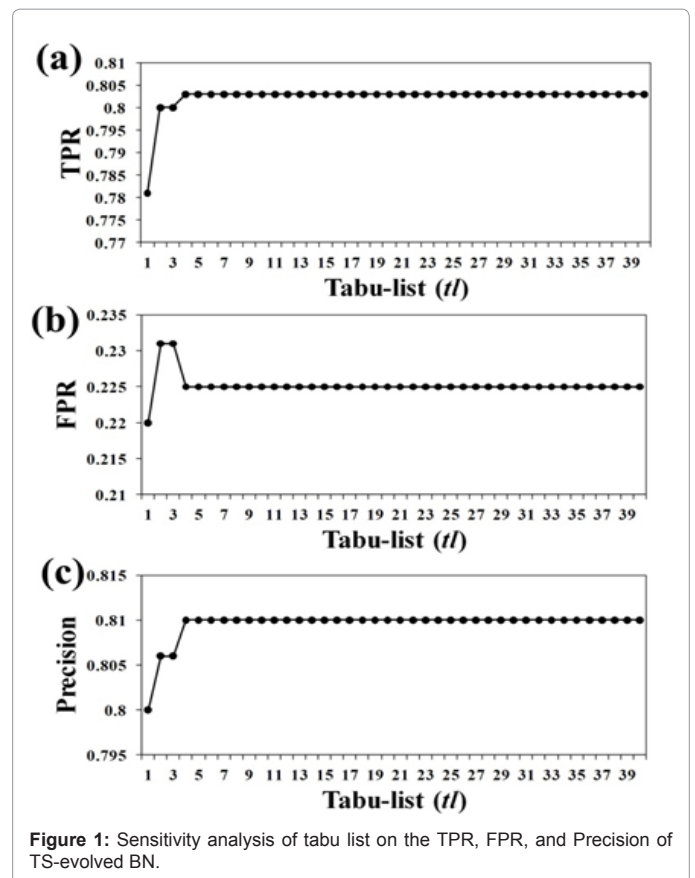


Figure 1: Sensitivity analysis of tabu list on the TPR, FPR, and Precision of TS-evolved BN.

accuracy, this study employed analyses of the Receiver Operating Characteristic (ROC) and the area under the ROC curve (AUC). The ROC curve is a popular tool in medical and imaging research. It conveniently displays diagnostic accuracy regarding sensitivity (i.e., TPR) against (1 – specificity) (i.e., FPR) at all possible threshold values [27,28]. By changing the cut off value, several different pairs of (1 – specificity) and sensitivity can be obtained [29]. In addition, the AUC can provide an indication of the accuracy of the test. If the ROC curve follows the diagonal, the diagnostic test is not useful and the area under this curve is 0.50, indicating that the test is not more specific or sensitive than random guessing [30].

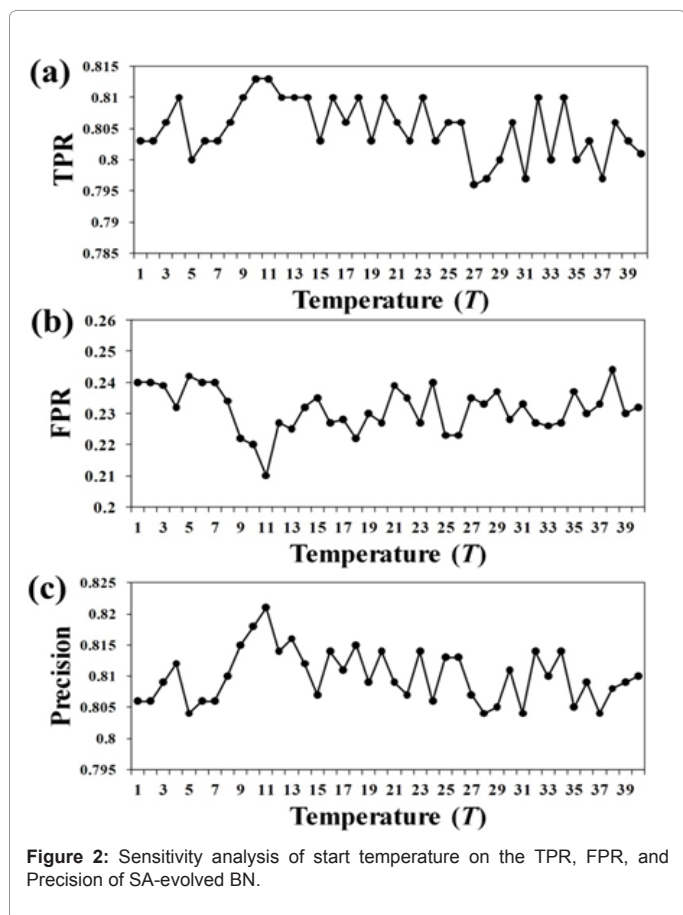


Figure 2: Sensitivity analysis of start temperature on the TPR, FPR, and Precision of SA-evolved BN.

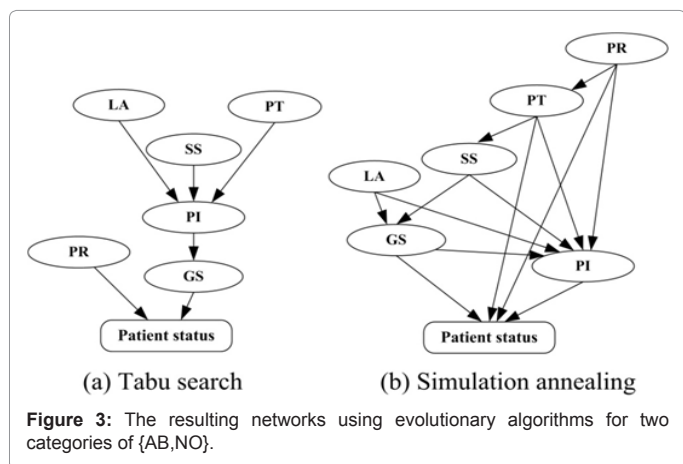


Figure 3: The resulting networks using evolutionary algorithms for two categories of {AB,NO}.

Network structures

First, for two categories, {AB, NO}, the networks that resulted from using TS are shown in (Figure 3a). In the figure, the attribute PR refers to parts of the network different from those of the other five attributes. This network exhibited two chains of causalities. The first chain shows that the patient status depends on the grade of SL, and the grade of SL in turn depends on the pelvic incidence and the other three attributes (i.e., pelvic tilt, lumbar lordosis angle, and sacral slope). Another edge in the network was located between patient status and pelvic radius. (Figure 3b) shows causality networks drawn using SA. The network exhibited more complex causality relationships among attributes than did TS. Similar to the structures formed when using two categories, for three categories, {DH, SL, NO}, the TS algorithm produced simple structures (Figure 4a), whereas SA yielded complex structures (Figure 4b).

Results of ROC and AUC analyses

The ROC curves obtained when using TS and SA for two and three categories are provided in (Figures 5 and 6), respectively. In this experiment, the AUC measures in each class, which summarize the statistic of diagnostic performance, are displayed in (Figure 7). According to the figure, higher weighted average AUC measures of classification accuracy were exhibited by the SA-evolved BN than by the TS-evolved BN, likely because SA develops more favorably than TS does, but yields complex network structures; therefore, both algorithms have high overall AUC measures. Although TS exhibited inferior AUC outcomes compared with those produced by SA, it is attractive for its simplicity and the few parameters that are necessary [20, 21].

Conclusion

In this paper, two heuristic structure learning algorithms used for probabilistic graphical models, namely TS and SA, were empirically evaluated and compared regarding efficiency. This study applied these algorithms in classifying the pathologies of the vertebral column. A data set containing values for six biomechanical features was used to classify patients into three categories, {DH, SL, NO}, and two categories, {AB, NO}. The results showed that using SA enhanced the AUC measures for predictions. For two categories, the average AUC values of 0.860 and 0.876 for TS and SA, respectively, were obtained. For three categories, the average AUC measures of 0.915 and 0.926 were obtained for TS and SA, respectively. The results indicate that SA is a more effective algorithm than TS. However, the empirical results obtained using TS indicated that the TS algorithm is promising because of its relatively simple network structure.

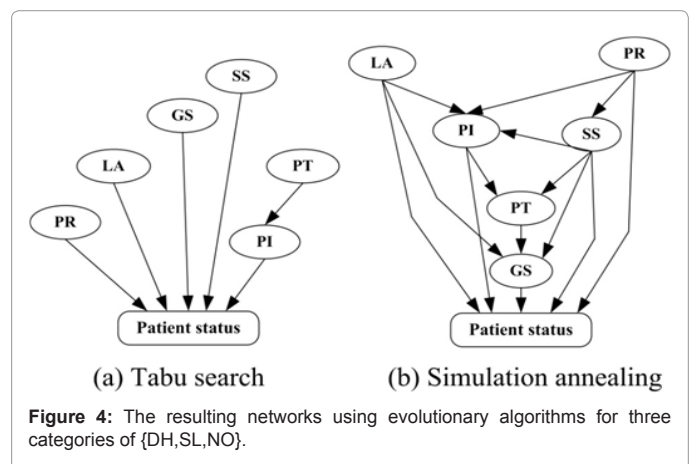


Figure 4: The resulting networks using evolutionary algorithms for three categories of {DH,SL,NO}.

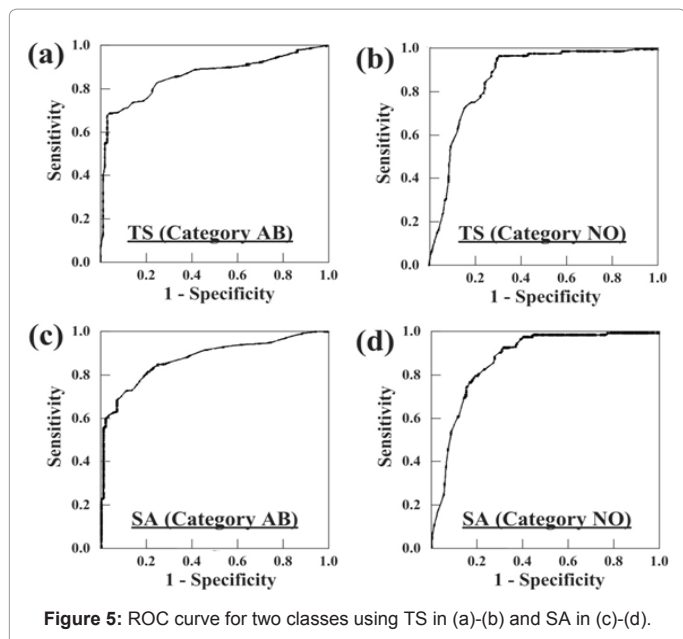


Figure 5: ROC curve for two classes using TS in (a)-(b) and SA in (c)-(d).

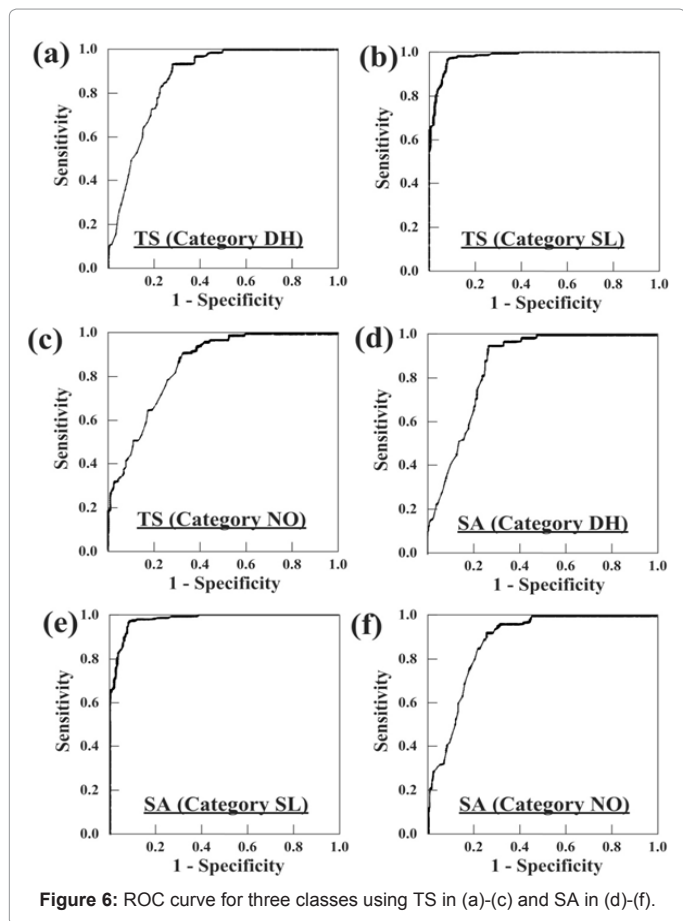


Figure 6: ROC curve for three classes using TS in (a)-(c) and SA in (d)-(f).

Acknowledgements

The support provided under grant no. NSC102-2313-B-464-001 by the National Science Council, Taiwan is greatly appreciated. The author is also grateful to the UC Irvine (UCI) machine learning repository for providing the data for the experiment.

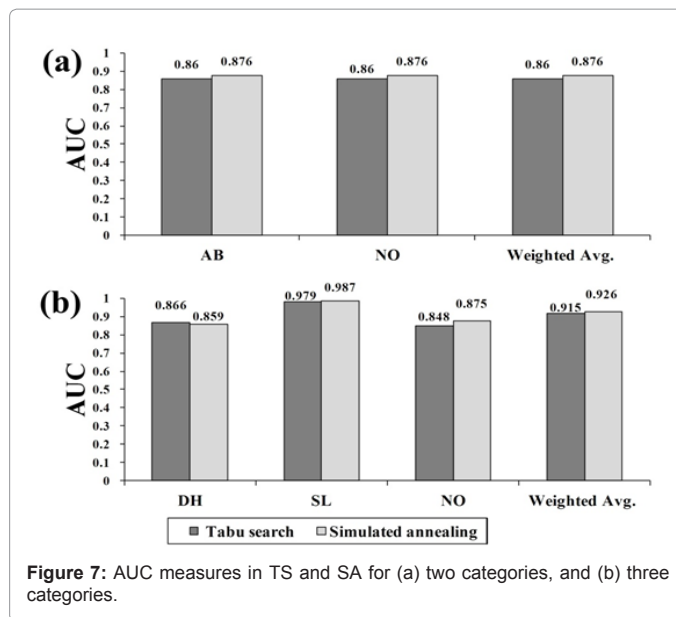


Figure 7: AUC measures in TS and SA for (a) two categories, and (b) three categories.

References

1. Onisko A, Druzdzal MJ, Wasyluk H (1999) A Bayesian Network Model for Diagnosis of Liver Disorders. Research Report CBMI-99-27, Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA.
2. Berry MJA, Linoff G (1997) Data Mining Techniques: For Marketing, Sales, and Customer Support. John Wiley and Sons, USA.
3. Bai X, Padman R (2005) Tabu Search Enhanced Markov Blanket Classifier for High Dimensional Data Sets. *Operations Research-Computer Science Interfaces Series* 29:337-354.
4. Cooper GF, Aliferis CF, Ambrosino R, Aronish J, Buchanan BG, et al. (1997) An Evaluation of Machine-Learning Methods for Predicting Pneumonia Mortality. *Artif Intell Med* 9: 107-138.
5. Pearl J (1998) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, San Francisco, USA.
6. Inza I, Larranaga P, Sierra B (2001) Feature Subset Selection by Bayesian Networks: A Comparison with Genetic and Sequential Algorithms. *Int J Approx Reason* 27: 143-164.
7. Chickering DM, Geiger D, Heckerman D (1994) Learning Bayesian networks is NP-hard. Technical Report MSR-TR-94-17, Microsoft Research, Redmond, WA.
8. Larranaga P, Kuijpers CMH, Murga RH, Yurramendi Y (1996) Learning Bayesian network structures by searching for the best ordering with genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics - Part A* 26: 487-493.
9. Wong ML, Lam W, Leung KS (1999) Using evolutionary programming and minimum description length principle for data mining of Bayesian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21: 174-178.
10. Jeon YJ, Kim JC, Kim JO, Shin JR, Lee KY (2002) An efficient simulated annealing algorithm for network reconfiguration in large-scale distribution systems. *IEEE Transactions on Power Delivery* 17: 1070-1078.
11. Wu H, Wang WP, Yang F (2011) Structure Learning Method of Bayesian Network with Uncertain Prior Information. *Energy Procedia* 13: 5421-5430.
12. Glover F (1986) Future Paths for Integer Programming and Links to Artificial Intelligence. *Comput Oper Res* 13: 533-549.
13. Tospornsampan J, Kita I, Ishii M, Kitamura Y (2007) Split-Pipe Design of Water Distribution Networks Using a Combination of Tabu Search and Genetic Algorithm. *World Academy of Science, Engineering and Technology*.
14. Frank A, Asuncion A (2013) UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine.
15. Langley P, Iba W, Thompson K (1992) An Analysis of Bayesian Classifiers. *AAAI-92 Proceedings*.

16. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian Network Classifiers. *Machine Learning* 29: 131-163.
17. Pernkopf F, O'Leary P (2003) Floating Search Algorithm for Structure Learning of Bayesian Network Classifiers. *Pattern Recogn Lett* 24: 2839-2848.
18. Bouckaert RR (1995) Bayesian Belief Networks: From Construction to Inference. University of Utrecht.
19. Glover F, McMillan C, Novick B (1985) Interactive Decision Software and Computer Graphics for Architectural and Space Planning. *Annals of Operations Research* 5: 557-573.
20. Hertz A, de Werra D (1987) Using Tabu Search Techniques for Graph Coloring. *Comput* 39: 345-351.
21. Taillard E (1991) Robust taboo Search for the Quadratic Assignment Problem. *Parallel Comput* 17: 443-455.
22. Kirkpatrick S, Gelatt JCD, Vecchi MP (1983) Optimization by Simulated Annealing. *Science* 220: 671-680.
23. Cerny V (1985) A Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm. *J Optimiz Theory App* 45: 41-51.
24. Xiaobo Z, Jiewen Z, Povey MJW, Holmes M, Hanpin M (2010) Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta* 667: 14-32.
25. Kalivas JH, Roberts N, Sutter JM (1989) Global Optimization by Simulated Annealing with Wavelength Selection for Ultraviolet-Visible Spectrophotometry. *Anal Chem* 61: 2024-2030.
26. Swierenga H, Wulfert F, De Noord OE, De Weijer AP, Smilde AK, et al., (2000) Development of Robust Calibration Models in NIR Spectroscopic Applications. *Analytica Chimica Acta* 411: 121-135.
27. Zou KH (2001) Comparison of Correlated Receiver Operating Characteristic Curves Derived from Repeated Diagnostic Test Data. *Acad radiol* 8: 225-233.
28. Metz CE (2006) Receiver operating characteristic analysis: A tool for the quantitative evaluation of observer performance and imaging systems. *Journal of the American College of Radiology* 3: 413-422.
29. Han UK, Kim YH (1998) Determination of Class II and Class III Skeletal Patterns: Receiver Operating Characteristic (ROC) Analysis on Various Cephalometric Measurements. *Am J Orthod Dentofac* 113: 538-545.
30. Daya S (2006) Diagnostic test: Receiver operating characteristic (ROC) curve. *Evidence-based Obstetrics and Gynecology* 8: 3-4.