Intervention Learning of Local Causal Structure Based on Sensitivity Analysis

Junzhao Li
School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, china
Email: Ljunzhao@163.com

Hongliang Yao, Jian Chang, Shuai Fang and Jianguo Jiang
School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, china
Email: {dmicyhl, dmiccj, fangshuai, jgjianghf} @163.com

***Abstract*—As intervened edges are difficult to be determined when intervention method is used for learning the causal relationships of probability model, an active learning method (Structural Intervention Learning of Sensitivity Analysis –SILSA Algorithm) is proposed. SILSA algorithm obtains original network structure based on k2 algorithm, then uses junction tree algorithm to decompose original networks structure and takes local intervention learning in every clique of junction tree, which can decrease the searching extension of intervened edges. Causal Bayesian networks can be learned by Edge-based Interventions when intervened edges are selected. In order to get appropriate intervened edge, sensitivity analysis is used to select the important edge in SILSA algorithm. The efficient of selecting intervened edge is improved. Experimental results show that the effectiveness and performance of SILSA algorithm are better than intervened edges with choosing randomly and passive learning method.**

***Index Terms*—causal relationship, sensitivity analysis, intervention learning, junction tree**

I. Introduction

Causal Bayesian networks are extensions to Bayesian networks in the semantics, it uses a directed acyclic graph which represents the probability relations between variables and describes the uncertainty problem in the causal knowledge [1]. Causal relationship has an important role in the prediction and inference. Since causal relationships with other non-causal relationships are intertwined and easy to be intervened, which makes the learning of Causal Bayesian networks difficulty.

During the past decade, people have proposed many methods for learning Bayesian network structure deal with observational data, such as the SEM algorithm [2] and K2 algorithm[3]. These learning methods can only learn the dependencies of probability. These methods cannot distinguish equivalent networks, so it is difficult to determine Causal Bayesian networks. These methods cannot learn the causality effectively.

Manuscript received March 20, 2012; revised June 1, 2012.
Corresponding author: Junzhao Li

Observational data generally submits to the same distribution, and certain elements in the system subject to external disturbances and affect the data change of distribution. The change of the distribution data is a basis of the detection of causality, and the method of learning causal knowledge by intervention is called Intervention Learning (IL). In recent years, intervention learning is paid attention by researchers, such as, Cooper proposed a method of causal knowledge by combining observational data and interventional data, which intervened node is randomly selected[4]. Eaton first assumed that the interference node is known, then given a learning method that the interventional data using an instance of learning in each learning step[5]. Jianxin discussed local causal learning methods by intervention[6]. Kuehnert discussed learning methods of soft intervention[7].

The above situations are based on the intervention of node, and the choice of intervened node is still difficult. Intervention learning of Causal Bayesian networks can perform by Edge-based Interventions. The intervened edge is intervened nodes which are connected by the edge. However, the number of edges is much larger than the number of nodes in network, intervened edges are also difficult to select. In Edge-based Interventions, the conditional probability distribution of intervened edge has the sensitivity characteristics, we can select intervened edge by sensitivity analysis (SA) algorithm.

SA algorithm studies sensitivity reaction of model conditional probability (or target probability), when parameters or evidence value come about a small changes. SA algorithm can be a quantitative analysis for the model parameter and the importance of structure, and has been widely applied in many areas such as medicine, civil engineering, computer science[8]. Sensitivity analysis in Bayesian networks, the researchers also did the related work [9]. They discovered that Bayesian networks is very sensitive to the accuracy of the parameter probability, and demonstrated that sensitivity analysis is very effective to Bayesian networks. Sensitivity analysis of Bayesian networks is extended to multiple parameters [10]. Renooij conducted a sensitivity analysis on dynamic Bayesian networks [11].

For a Bayesian networks with n nodes, the number of

doi:10.4304/jcp.8.4.912-919

possible edges is $n(n-1)$. So intervened edge is difficult to select. Junction tree algorithm can reduce the complexity exponentially and not lose the joint probability[12]. As a result, Junction tree algorithm can be used to decompose Bayesian networks into cliques, then in the every clique Structural Intervention Learning can decrease the search range of intervention edge.

We introduce a new active learning algorithm for causal Bayesian networks with Edge-based Interventions and sensitivity analysis. Our objective is to learn the causal Bayesian network structure that achieves the specified structure accuracy with a minimal number of interventions. We obtain original network structure based on k2 algorithm. Sensitivity analysis is used to select intervened edge in cliques which is produced by taking junction tree algorithm to decompose original networks structure. Causal Bayesian networks can be learned by Edge-based Interventions. We examine the effectiveness and efficiency of the proposed method on identifying causal relationships based on three benchmark Bayesian networks and compare our method with some other major methods.

II. Construct And Decompose Original Networks

A. *K2 Algorithm*

K2 algorithm is a simple and efficient structure learning algorithm of Bayesian networks. The topological sequence of all variables is given at the beginning. K2 algorithm uses a greedy search method to construct the network structure. First, it initializes each node's parent as an empty set. Second, it tests possible parent node start from the second node in the sequence and computes the scoring function (such as the BIC and the MDL or Bayesian score). It adds the node with the maximum score function to the parent node set. These steps are repeated until the searching of all the node parent is end or achieves maximum father node number.

B. *Decompose Networks Based On The Correlation Of Variables*

Bayesian networks learning has a high time complexity and space complexity, it is difficult to learn the network structure effectively. It needs to decompose the Bayesian networks. The decomposability of Probabilistic graphical model and Junction tree algorithm can constitute cliques by variables which has strong correlation. The decomposability definition of the graph as follows[13].

Definition 1 (Moral graph): Moral graph refers to the graph that connects all the parent nodes which have common child with the non-directional edges. Operations of generating Moral graph called normalization.

Definition 2 (Graph with chord): Graph with chord G refers to a directed graph, any of the ring with a length greater than 3 has chord.

Definition 3 (Clique): Clique refers to the largest complete graph, the largest complete graph refers to all of the nodes in graph are connected. The complete graph in graph with chord which not be contained in other complete graphs is a clique.

Definition 4 (Decomposability of Probabilistic graphical model): The Decomposition tree of graph $G(V,E)$ is a tuple $D=(S,T)$, $S=\{X_c \mid c\in C\}$ is the set of Node subsets. $T=(C,F)$ is a tree, nodes in the tree are the elements in S and satisfy the following three conditions.

(1) $\bigcup_{c\in C} X_c = V$.

(2)Every edge $(v,w)\in E$ has a subset $X_c \in S$ contains v and w.

(3)For every node $v\in V$, the set of node $\{c \mid v\in X_c\}$ constitutes a connected subtree of T.

Definition 5 (Junction tree): A Junction tree is a tuple $T=(\Gamma,\Delta)$, Γ is the set of nodes in clique, two cliques in Γ are connected by clique nodes in Δ. For any pair of adjacent cliques, $C_i, C_j \in \Gamma$, $S_k \in \Delta$ is a Split Clique between C_i and C_j .

Junction tree algorithm is described as follows.

(1) Generate Moral graph. Moral graph is the graph which turns the directed edges in Directed acyclic graph to undirected edges, shown in Fig. 1(a). A Moral graph is shown in Fig. 1(b). Curved edges in graph are new edges.

(2) Triangulation of graph. For the ring which contains four and more nodes, a undirected edge is added to connect two non-adjacent nodes in ring. Fig. 1(c) is the result of the processing of the triangulation of Moral graph. The dotted line is the newly added edge.

(3) Identify of cliques. To determine the clique node in Triangulation graph, it needs to merge the cliques which have relation and calculation the number of nodes in every clique. Each clique is a complete subgraph of undirected graph.

For every clique $C_s \in \Gamma$, $|C_s|$ is the number of nodes in C_s. For a given positive integer threshold T (3, 4 or 5, according to the actual situation). The clique which satisfies $|C_s| \geq T$ and considers whether needs to merge the adjacent Cliques. In Fig. 1 (c) the number of cliques which contain more than 3 nodes : $C_1=\{X_2,X_4,X_5,X_6,X_7,X_8\}, C_2=\{X_1,X_2,X_4\}, C_3=\{X_2,X_3,X_7\}$.

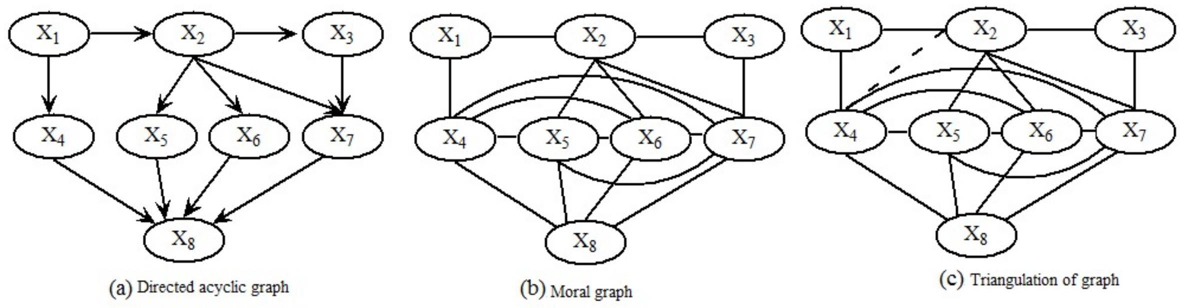

Fig.1 Clique formation

III. STRUCTURAL INTERVENTION LEARNING

Definition 6 (Structural intervention): Given a variable set V, S is subset of V. We have the following definition for a structural intervention I_S on a variable in S:

(1)There is no common causal between I_S and any variable of V. There is no reason variable of I_S in V.

(2)When $I_S = k$, and $K \neq 0$, I_S makes every variables of S independent of its causes. I_S determines the distribution of S, that is, in the factored joint distribution $P(V)$, $P(S \mid pa(S))$ is replaced with $P(S \mid I_S = k)$, all other terms in the factored joint distribution are unchanged, in which $pa(S)$ is the parents of S.

The definition of structural intervention indicates that the causal structure can be manipulated, and all edges between variables that are subject to an intervention are removed.

Given a causal Network $G = (V, E)$, and a variable $X \in V$, an intervention variable I is added to intervened variable X in G, represented by $I \rightarrow X$. The value of intervened variable is in $\{idle, do(x_i)\}$ when the value of x_i is in the range of X_i. Where *idle* shows there is no intervention, $do(x_i^{'})$ expresses that the intervened variable is forced to be $x_i^{'}$. Intervention forces intervened variable to take a specific value, and destroys the normal casual relationship of intervened variable. The distribution of X_i is conditional reflected by pa_i. The probability distribution equation of intervened variable is equation (1) [14] and Fig. 2 shows that intervention manipulative process.

Intervention of the single variable, Pearl proposed the do-calculus method to calculate the influence of intervention variables V. Let X,Y,Z be arbitrary disjoint sets of nodes in a causal graph G. We denote by $G_{\overline{X}}$ the graph obtained by deleting from G all arrows pointing to X .Likewise, we denote by $G_{\underline{X}}$ the graph obtained by deleting from G all arrows emerging from nodes in X. To represent the deletion of both incoming and outgoing arrows, we use the notation $G_{\overline{X}\underline{Z}}$. The expression $P(y \mid \hat{x}, z) \stackrel{\Delta}{=} P(y, z \mid \hat{x}) / P(z \mid \hat{x})$ stands for the probability of $Y = y$ given that X is held constant at x and that (under this condition) $Z = z$ is observed.

$$P(X_i \mid pq, I_i) = \begin{cases} P(X_i = x_i \mid pq) & if\ I_i = idle \\ 0 & if\ I_i = do(x_i^{'})\ and\ x_i \neq x_i^{'} \\ 1 & if\ I_i = do(x_i^{'})\ and\ x_i = x_i^{'} \end{cases} \quad (1)$$

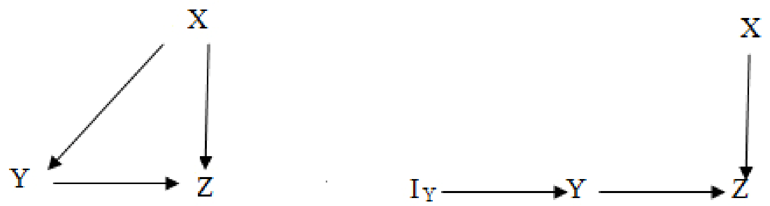

Fig.2 Structural Intervention

The do-calculus rules: Let G be the directed acyclic graph associated with a causal model. Let $P(.)$ stand for the probability distribution induced by that model. For any disjoint subsets of variables set X,Y,Z and W we have the following rules.

(1) Insertion/deletion of the observations

$$P(y \mid \hat{x}, z, w) = P(y \mid \hat{x}, w) \text{ if } (Y \perp Z \mid X, W)_{G_{\overline{X}}} \quad (2)$$

(2) Action /observation exchange

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, z, w) \text{if } (Y \perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}} \quad (3)$$

(3) Insertion/ deletion of actions

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, w) \text{ if } (Y \perp Z \mid X, W)_{G_{\overline{X}\overline{Z(W)}}} \quad (4)$$

$Z(W)$ is the set of Z-nodes that are not ancestor of any W-node in $G_{\overline{X}}$. The do-calculus rules are complete[15], which was proved. Pearl took this rule to specify conditions for the identification of causality. The casual structure changes with the change of Joint probability distribution of variables. Once structure intervention may operates many variables by same way. In order to get the external influence of causal structure, structural intervention requirements distribution function of intervened variables can be determined by the intervention.

IV. Choosing The Intervention Edges Based On Sensitivity Analysis

Choosing intervention edges randomly will lead to the uncertainty of choice, which affects the learning results. it is necessary to select the intervention edges under a certain standard. We can calculate the importance of each parameter by sensitivity analysis method [16]. The important edges are selected based on the importance of parameters, and the important edges are used for intervention edges. Since the parameters and edges are correspond to each other, such as $P(x_i \mid pa(x_i))$ .There is an edge between the node x_i and its parents $pa(x_i)$, so we can analyze the importance of edges in clique by the importance of parameters.

Definition 7 (Sensitivity function): Given the parameter $\theta = P(H = h_i \mid pa(H))$, Where h_i is a value of a variable H, $pa(H)$ is the set of H's parents. $P(B = b \mid E = e)$ represents the posterior probability when the value of E is e and the value of B is b, short for $P(b \mid e)$. The sensitivity of the θ for target probability $P(b \mid e)$ can be represented by sensitivity function $f_{P(b|e)}(\theta)$.

We have the following definition for Sensitivity function [17].

$$f_{P(b|e)}(\theta) = \frac{P(b,e)(\theta)}{P(e)(\theta)} = \frac{\alpha\theta + \beta}{\gamma\theta + \sigma} \quad (5)$$

$\alpha, \beta, \gamma, \sigma$ are constants in Eq. (5), $P(b,e)(\theta)$ is the joint probability distribution of b and e about θ , $P(e)(\theta)$ is the prior probability of e about θ .

Definition 8 (Parameter sensitivity):For Bayesian networks, θ is a probability parameter $\theta = P(h_i \mid pa(H))$, y is the target probability $y = p(b \mid e)$, e is a evidence added into the network. For partial derivative of θ in Eq. (6).

$$S(\theta \mid y, e) = \frac{\partial p(y \mid e)}{\partial \theta} \quad (6)$$

According to the parameter sensitivity definition, we can determine what parameters is important in the network, the parameters with high sensitivity will influence the generating of target probability more.

Definition 9 (Importance of Parameters): For Bayesian Networks, θ is a probability parameter $\theta = P(h_i \mid pa(H))$, y is the target probability $y = p(b \mid e)$. The importance of parameters θ is the average value of sensitivity where the y and e have proper value.

$$I(\theta) = \frac{1}{rs} \sum_{y,e} S(\theta \mid y, e) = \frac{1}{rs} \sum_{y,e} \frac{\partial p(y \mid e)}{\partial \theta} \quad (7)$$

Definition 10 (Importance of Edges): Given the nodes A and B in clique, $\theta_{ij} = p(A = a_i \mid B = b_j)$, a_i, b_j is the value of A and B. m and n is the number of attribute value of A and B. The importance of edge $B \to A$ as follows,

$$EI(B \to A) = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} I(\theta_{ij}) \quad (8)$$

Where the $EI(B \to A)$ represents the importance of $B \to A$, $I(\theta_{ij})$ is the importance of $P(A \mid B)$ in clique.

Intervened edges have are more importance in clique. The edge which has the most importance is defined in Eq. (9).

$$\begin{aligned} EI_{\max}(B \to A) &= \max_{k=1}^{N}(EI(B \to A)) \\ &= \max_{k=1}^{N}(\frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} I(\theta_{ij})) \end{aligned} \quad (9)$$

Where n is the number of edges in clique (we do not consider the edge which added in the generating process).

Based on the Parameter sensitivity, we can calculate the importance of every parameter in cliques by Eq. (7). Then calculate the importance of each edge by Eq. (8), We can discover the most important edge by Eq. (9), and take the most importance edge as the intervened edge in cliques.

V. SILSA Algorithm

Since the direction of edges in every clique has nothing to do with other cliques, each clique is the basic unit.

Given a clique, we can determine the importance of edge by sensitivity analysis on parameters in this clique. We can discover the direction and causal of edges by Edge-based Interventions when intervened edges are selected.

A. Criteria For Edge Orientation

One edge is chosen with certain criterion for an edge-based intervention, and the edge-based intervention determines whether there is a causal influence relationship from the parent node to the child node in the selected edge. We apply Koivisto's exact method [18]to estimate the edge probabilities with the available data and topological constraints. The edges are predicted as the learned edges when their probabilities are greater than 0.5. This method utilized the fact that the order of the parents of a variable is irrelevant to the variable's probability estimation, but the exact method can be applied to domains with a moderate number of variables.

B. Stop Criteria For Causal Structure Learning

For every pair of variables, three possible situations between them are usually considered: an edge from A to B ($A \rightarrow B$), an edge from B to A ($B \rightarrow A$), or no edge between A and B ($A \perp B$). The probabilities of the edges given the available data D and domain knowledge K can be written as:

$$P(A \rightarrow B \mid D, K) = \sum_{A \rightarrow B \in E(G)} P(G \mid D, K) \qquad (10)$$

where $P(G \mid D, K)$ is the probability of Bayesian networks G given the data D and
domain knowledge K, and E(G) is the set of edges in Bayesian networks G.

The edge entropy of structure $A \rightarrow B$ is defined as follows:

$$\begin{aligned} \mathrm{H}_S(A,B) &= -p(A \rightarrow B)\log p(A \rightarrow B) - \\ & p(A \leftarrow B)\log p(A \leftarrow B) - p(A \perp B)\log p(A \perp B) \end{aligned} \qquad (11)$$

So the entropy of Bayesian networks G [19] can be written as:

$$\mathrm{H}s(G) = \sum_{A,B} \mathrm{H}s(A,B) \qquad (12)$$

In an ideal condition, only one of the three conditions between variables A and B is with probability 1 and other two are with probability 0. The entropy between variables A and B will be 0. If all pairs of variables are ideal, the entropy of the real DAG will be 0.

When to stop the learning process in Bayesian network learning for causal knowledge discovery in practice must be designed carefully. One way is to choose a fixed number of interventions as the stop criterion. The disadvantage of this approach is that there is no guarantee on the quality of the learned Bayesian network structure. In experiment we use average entropy of Bayesian networks and average Hamming distance as the stop criterion. The ideal entropy and Hamming distance of the learned structure are 0. We give different threshold values of entropy and Hamming distance as the stop criteria.

C. The Experiment Method

The experiment setup is as follows:

(1) Choose one Bayesian networks from the three Bayesian networks as the ground truth Bayesian networks. Sample an observational data set with n instances from the ground truth Bayesian network;

(2) k2 algorithm is used to obtain original networks from observational data, Junction tree algorithm is used to discover cliques.

(3) Sensitivity analysis is used to compute the importance of edges as intervened edges in all cliques.

(4) Estimate the edge probabilities to predict the orientation of which selected from every clique. Perform random edge selection for intervention, or other learning algorithms.

(5) Check the stop criterion. If the stop criterion is satisfied, stop. otherwise, Generate a set of new interventional data with m instances from the ground truth, combine the new data with the existing data as the new available data. return to step(2).

VI. Experimental Results

The proposed method has been tested in experiments with the benchmark Bayesian networks. The three benchmark Bayesian networks are Asia network, car network and cancer network. Asia network is a Chest-clinic network which has 8 nodes (each with 2 values) and 8 arcs. There are 5 variables in Cancer network and 12 variables in Car network. We conducted the simulations under MATLAB (version 7) with the support of two tools, one is Bayesian networks toolbox(BNT) [20]and the other is BDAGL[21].

In this experiment, we compared the learned structure with the ground truth Bayesian networks. Two different situations are tested in our experiments which are the average hamming distance and the structure entropy of the learned Bayesian networks. Different numbers of interventions have been tested in our experiments. The maximal number of total interventional instances is set to 2000 for Asia network, 300 for Cancer network and 1000 for Car network.

We generate data from benchmark Bayesian networks and carry through experiment through combining observational data and intervention data.

For evaluate the accuracy we introduce sensitivity which is calculated in Reference [22]. Sensitivity is the number of correctly identified edges over the total number of edges. Table 1 shows the comparison of SILSA algorithm and the other two algorithms. GES(Greedy Equivalent Search) and TPDA[23]are effective Bayesian network structure learning methods deal with observational data at present. In Table 1 we showed that SILSA algorithm outperforms on average other two algorithms in terms of sensitivity

TABLE I.
RESULT ON DIFFERENT NUMBER OF ASIA DATA

algorithm	case number	true edge	reverse edge	redundance edge	lost edge	sensitivity
TPDA	500	2	1	3	5	25%
	1000	3	1	3	4	37.5%
	2000	3	1	3	4	37.5%
GES	500	3	4	1	1	37.5%
	1000	6	2	1	0	75%
	2000	7	0	2	1	87.5%
SILSA in paper	500	7	1	0	1	87.5%
	1000	7	0	0	1	87.5%
	2000	8	0	0	0	100%

Next we tried to learn back the car structure using different cases which contains observational samples and interventional samples and Compare with real network. In this experiment, we compared the learned structure with the ground truth Bayesian networks. The difference between the learned structure and the ground truth is measured with average Hamming distance. In Fig. 3, the lines represent the change of the average Hamming distance with the number of interventions. The initial instances to be collected are set to 400, the size of the interventional data in each active learning step increases 40. Fig. 3 shows SILSA algorithm that we proposes in paper leads to Bayesian networks with the smallest average Hamming distance to the ground truth, as compared with other methods.

Fig. 4 shows the changes of the average Hamming distance with the number of interventions when the initial instances are set to 1000, the size of the interventional data in each active learning step increases 100. Fig. 5 shows the situation when the initial instances are set to 2000 and the interventional data increases 200 in each step. With more interventional data, the average distance from the learned structure to the ground truth will be smaller. With 11 interventions, the average distance of SILSA algorithm is 0 in Fig.4. The average distance of SILSA algorithm reaches 0 with less interventions when the initial instances are 2000 in Fig.5. The variance of the Hamming distance from SILSA algorithm is the lowest, while the variances of the Hamming distances from GES and random intervention are high.

Fig.6 shows that horizontal ordinate is the number of intervention and vertical coordinates is the average structural entropy compare real network with network study. The structure learned with observational data has the highest entropy. The initial instances are set to 300, the size of the interventional data in each active learning step increases 30. The entropy of Bayesian network structure learned with SILSA algorithm is minimum compare with others, it illustrates that SILSA algorithm is more effective.

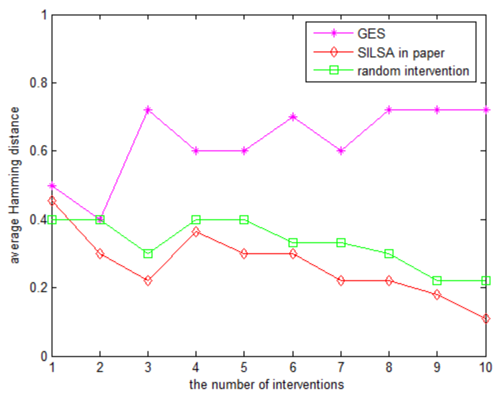

Fig.3 Relationship between the number of interventions and the average Hamming distance from the learned Bayesian network structure to the ground truth from car network when initial instances are set to 400

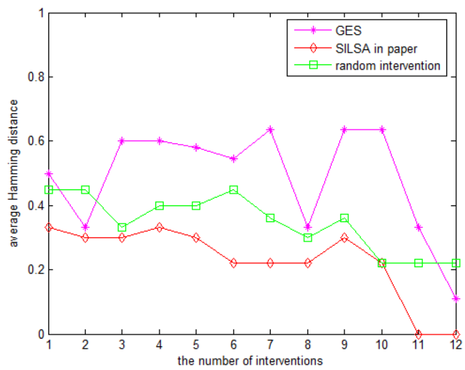

Fig.4 Relationship between the number of interventions and the average Hamming distance from the learned Bayesian network structure to the ground truth from car network when initial instances are set to 1000

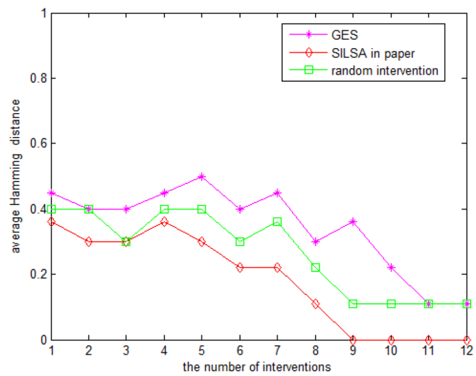

Fig.5 Relationship between the number of interventions and the average Hamming distance from the learned Bayesian network structure to the ground truth from car network when initial instances are set to 2000

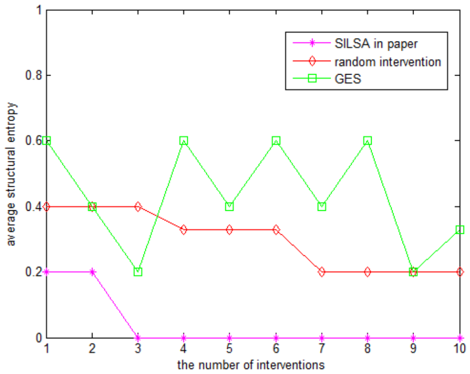

Fig.6 Relationship between the number of interventions and the structure entropy of the learned Bayesian network from Cancer network

VII. CONCLUSIONS

In this paper we presented a new algorithm for Bayesian network structure learning, called Structural Intervention Learning of Sensitivity Analysis (SILSA) algorithm. The algorithm combines ideas from intervention learning, sensitivity analysis, and junction tree method in a principled and effective way. It first learns the skeleton of a Bayesian network from observational data and then performs junction tree method to get clique, and orient the edges based on sensitivity analysis and intervention learning.

Experiments show that SILSA algorithm can reach the required structure entropy and hamming distance with smaller number of interventions than random intervention, and much better than merely estimating the structure with observational data. The statistical test also shows that SILSA algorithm is an effective method for learning causal Bayesian network structure when initial instances are small.

ACKNOWLEDGEMEN

This work is supported by the National Natural Science Foundation of China (Grant No. 61070131, 61175051, 61175033)

REFERENCES

[1] David Heckerman, A Tutorial on Learning with Bayesian Networks, in Learning in Graphical Models, Jordan, Editor, MIT Press, Cambridge, 1998, pp.301-354.

[2] Shuangcheng Wang, Senmiao Yuan, "Research on Learning Bayesian Networks Structure with Missing Data", Journal of Software (in Chinese), Vol. 15, No.7, 2004,pp.1042-1048.

[3] Gregory F. Cooper, Edward Herskovits, "A Bayesian method for the induction of Probabilistic Networks from data", Machine Learning, Vol. 9, No.4, 1992,pp.309-347.

[4] Gregory F. Cooper, Changwon Yoo, "Causal Discovery from a Mixture of Experimental and Observational Data", Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence(UAI), San Francisco, CA, 1999,pp.116-125.

[5] Daniel Eaton , Kevin Murphy, "Exact Bayesian structure learning from uncertain interventions", International Conference on Artificial Intelligence and Statistics, San Juan, Puerto Rico,2007,pp.107–114.

[6] Jianxin Yin, You Zhou, Changzhang Wang, Ping He , Cheng Zheng, Zhi Geng, "Partial orientation and local structural learning of causal networks for prediction", JMLR W&CP, WCCI2008 workshop on causality, Hong Kong,2008,pp.93-105.

[7] Christian Kühnert, Thomas Bernard, Christian Frey, "Causal structure learning in process engineering using Bayes Nets and soft interventions", Industrial Informatics (INDIN), 2011 9th IEEE International Conference on, Lisbon, Portugal, 2011,pp.69-74.

[8] Enrique Castillo, Roberto Mínguez, Carmen Castillo, "Sensitivity Analysis in Optimization and Reliability Problems", Reliability Engineering and System Safety, Vol. 93, No.12, 2008,pp.1788-1800

[9] Haiqin Wang, Building Bayesian Networks: Elicitation, Evaluation, and Learning, Ph.D dissertation, University of Pittsburgh, Pittsburgh, 2004.

[10] Hei Chan, Adnan Darwiche, "Sensitivity Analysis in Markov Networks", Nineteenth IJCAI, Edinburgh, Scotland, 2005,pp. 1300-1305.

[11] Silja Renooij, "Efficient Sensitivity Analysis in Hidden Markov Models", 2010,http://www.helsinki.fi/pgm2010/papers/renooij1.pdf

[12] S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their applications to expert systems", Journal of the Royal Statistical Society, Vol. 50, No.2, 1988, pp.154-227.

[13] Uffe Kjaerulff, "Reduction of Computational Complexity in Bayesian Networks through Removal of Weak Dependences", Proc of the 10th Annual Conference on Uncertainty in Artificial Intelligence, Seattle,USA,1994,pp.374-382.

[14] Judea Pearl, Causality: Models, Reasoning, and Inference, New York, Cambridge University Press, 2000.

[15] Yimin Huang, Marco Valtorta, "Pearl's calculus of intervention is complete", Proceedings of the 22nd Conference on Uncertainty and Artificial Intelligence, Corvallis, Oregon: AUAI Press, 2006,pp.437-444.

[16] Haiqin Wang, Irina Rish, Sheng Ma, "Using sensitivity analysis for selective parameter update in Bayesian network learning", Information Refinement and Revision for Decision Making: Modeling for Diagnostics, Prognostics and Prediction, AAAI 2002 Spring Symposium, Menlo Park, CA:AAAI Press, 2002,pp.29–36.

[17] Veerle M.H. Coupé , Linda C. van der Gaag, "Properties of sensitivity analysis of Bayesian belief networks", Annals of Mathematics and Artificial Intelligence, Vol. 36, 2002 ,pp.323–356.

[18] Mikko Koivisto, "Advances in exact Bayesian structure discovery in Bayesian networks" , Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, Menlo Park, CA:AUAI Press,2006,pp.241–248.

[19] Simon Tong, Daphne Koller, "Active Learning for Structure in Bayesian Networks", Proceedings of international joint conference on Artificial intelligence, Seattle, WA,USA,2001,pp.863–869.

[20] Kevin Murphy, Bayesian Net Toolbox for Matlab (2007). http://code.google.com/p/bnt/

[21] Daniel Eaton, Kevin Murphy, BDAGL: Bayesian DAG learning. (2007). www.cs.ubc.ca / ~murphyk /Software/BDAGL/

[22] Ioannis Tsamardinos, Laura E. Brown, Constantin F. Aliferis, "The max-min hill-climbing Bayesian network

structure learning algorithm" ,Machine Learning, Vol. 65, No.1, 2006,pp.31-78.
[23] Jie Cheng, Russell Greiner, Jonathan Kelly, David Bell b, Weiru Liu, "Learning Bayesian networks from data: An information-theory based approach", Artificial Intelligence, 2002, pp.43-90.

Junzhao Li was born in AnHui Province, China, in 1975. He received the M.S. degrees in Graduate University of Chinese Academy of Sciences. He is currently a Ph.D. candidate in Computer Science in Hefei University of Technology. His research fields include machine learning and artificial intelligence.
He is a lecturer of Hefei University of Technology.

Hongliang Yao was born in AnHui Province, China, in 1975. He received Ph.D. degrees in Hefei University of Technology.
He is an associate professor of Hefei University of Technology.

Jian Chang was born in AnHui Province, China, in 1983. He received the M.S. degrees in Hefei University of Technology.

Shuai Fang was born in AnHui Province, China, in 1978, She received the Ph.D. degrees in Harbin Institute of Technology.
She is an associate professor of Hefei University of Technology.

Jianguo Jiang was born in AnHui Province, China, in 1955. He is a professor of Hefei University of Technology.