Knowledge-Based Systems 50 (2013) 87–99

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Ranking evaluation of institutions based on a Bayesian network having a latent variable

Jun-Seong Kim, Chi-Hyuck Jun *

Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea

ARTICLE INFO

Article history:
Received 5 April 2012
Received in revised form 29 April 2013
Accepted 22 May 2013
Available online 31 May 2013

Keywords:
Ranking estimation
Linear Gaussian model
Structure learning
Gibbs sampling
Multiple search
Causal discovery

ABSTRACT

This paper proposes a new probabilistic graphical model which contains an unobservable latent variable that affects all other observable variables, and the proposed model is applied to ranking evaluation of institutions using a set of performance indicators. Linear Gaussian models are used to express the causal relationship among variables. The proposed iterative method uses a combined causal discovery algorithm of score-based and constraint-based methods to find the network structure, while Gibbs sampling and regression analysis are conducted to estimate the parameters. The latent variable representing ranking scores of institutions is estimated, and the rankings are determined by comparing the estimated scores. The interval estimate of the ranking of an institution is finally obtained from a repetitive procedure. The proposed procedure was applied to a real data set as well as artificial data sets.

1. Introduction

Ranking higher education institutions began in the United States in the early 20th century, and it has become increasingly more influential with time. Most ranking systems select the performance indicators (research, education, etc.), assign a weight to each indicator and calculate the weighted score for each institution; this is the 'weight-and-sum' approach. However, this method sometimes becomes controversial because assigning weights to performance indicators is subjective. Thus demand for a new quantitative method of ranking is on the rise although there are many issues other than raking methods as pointed out in Vught and Westerheijden [34].

Some studies have used approaches other than the weight-and-sum approach to rank institutions. These methods include an empirical Bayes approach to ranking schools based on student achievement [18], Bayesian analysis for ranking institutions [15], a latent-variable technique for university ranking based on several indicators [16] and the analytic hierarchy process for determining weights of indicators [23]. In addition, a method based on a supervised naïve Bayes structure uses the mixture of truncated exponentials, which applies the rank information of an expert [12]. A Bayesian latent variable model has been proposed for estimating the top ranked SNPs detected from genetic association studies [13], which may also be applied to the institution ranking problem.

When using the ranking evaluation systems other than the weight-and-sum approach, a structure analysis is necessary which considers the relationship among performance indicators. However, few studies have been conducted on the structure analysis for ranking evaluation. The ordinary modeling techniques for the structure analysis are structural equation models (SEMs) based either on linear structure relationships or partial least squares and Bayesian networks (BNs) [20]. Especially, BNs can be used as probabilistic inference engines, building models of domains that have intrinsic uncertainty. BNs are graphical models based on the notion of conditional independence that subsumes a wide range of statistical models including regression models, factor analysis models, and structural equation models. In a BN, a directed acyclic graph (DAG) represents a set of conditional independence constraints among a given number of variables and their related conditional probability distributions. The procedures for developing BNs involve learning the structure (the relationships between variables) first, and then parameterizing the associated conditional distributions. Graphical models, in particular those based on DAGs, have natural causal interpretations and thus form a language in which causal concepts can be discussed and analyzed in precise terms [21]. The conditional independence assumptions in a BN yield models more compact than those based on full joint probability distributions, thus reducing computational complexity when the number of variables is large [30]. Lately, BNs have been used in various applications such as risk management [3], resource allocation decisions [10,11], IT implementation [20], species distribution [1], higher education [12], and health risk assessment [22].

* Corresponding author. Address: Industrial and Management Engineering, POSTECH, Republic of Korea. Tel.: +82 542792197.
E-mail addresses: soulloan@postech.ac.kr (J.-S. Kim), chjun@postech.ac.kr (C.-H. Jun).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.05.010

In this paper, we propose a new Bayesian network model which has an unobservable latent variable that affects all other observable variables. To solve the ranking problem, the latent variable represents the ranking variable to be finally estimated, and the observable variables indicate performance indicators of each institution. Our first task is to identify the causal relation among these variables, and the second task is to determine the institution rankings in terms of intervals.

Causal discovery algorithms (CDAs) for a BN can be classified into three different categories: the score-based approach, the constraint-based approach, and the combined approach [35]. Score-based methods select a model with the highest posterior probability when the prior of each model is given. Constraint-based methods use statistical methods to detect associations and independencies among variables. Each method has a trade-off between time complexity and accuracy. Score-based methods generally provide accuracy but their time complexity is very high. Constraint-based methods give much lower time complexity, but they may not be accurate if conditional independence (CI) tests fail. The combined methods use both concepts to build the graph by compromising the benefit in terms of time complexity and accuracy. These methods cannot be directly applied to our new graphical model which has a latent variable, so a new approach needs to be developed for our purpose.

The proposed approach consists of two phases and each phase is repeated until the network structure converges. First, latent rankings are obtained using Gibbs sampling and gradient descent from a given network structure. Second, the updated network structure is found by the revised version of Multiple Search (MS) algorithm [7] using the latent rankings. The final rankings in terms of interval estimates are obtained based on the convergent structure.

The rest of the paper is organized as follows. The BN under consideration is modeled in Section 2. The proposed method of learning the BN is described in Section 3. The proposed method is applied to real data in Section 4. Section 5 reports on a further experiment with artificial data to check the accuracy and consistency of the proposed new Bayesian network model. Conclusions are presented in Section 6.

2. Bayesian network having a latent variable

Our ranking problem is to assign a ranking to each of the given number of institutions based only on their performance indicators. For this purpose we will consider a new type of BN as shown in Fig. 1. The difference from the usual BN is to contain an unobservable latent variable that affects all other observable variables. Let X_i ($i = 1,2,\ldots,m$) be the ith variable where m is the number of variables. These variables are performance indicators for our ranking problem. They are assumed to be observable for each institution, and they may have causal relationships among themselves. Let Z be an unobservable latent variable that affects all other variables. Without the latent variable Z, the network will be the same as a usual graphical model. For our ranking problem, Z_j ($j = 1,2,\ldots,n$) represents the unobservable ranking score of the jth institution, where n is the number of institutions. Hence, our problem is to estimate the ranking scores of all institutions by considering the causal relationship among observable performance indicators. Once the ranking scores of all institutions are estimated, the ranking will be determined by comparing the scores' magnitudes.

2.1. Linear Gaussian model

Consider an arbitrary DAG in which node i represents a continuous random variable X_i that has a Gaussian distribution. Then, under the linear Gaussian model, the conditional distribution of X_i given its parent node pa_i is also Gaussian:

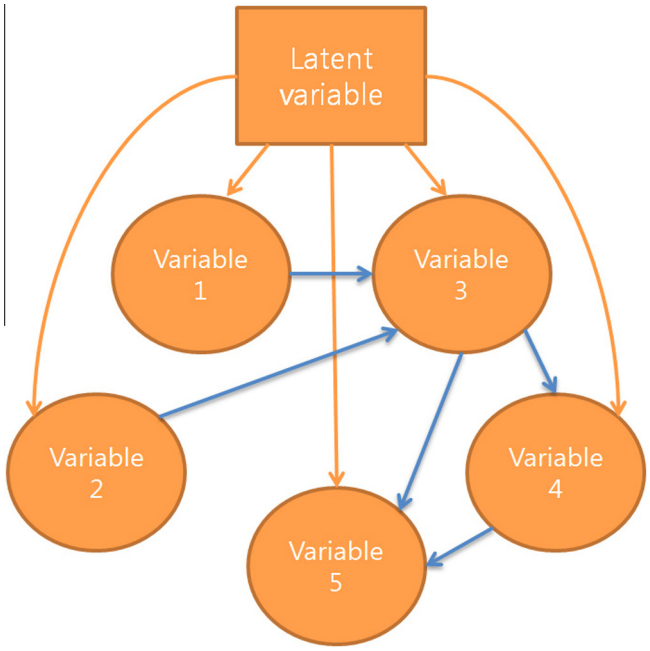

Fig. 1. An example of our graphical model having a latent variable.

$$X_i|X_{pa_i} \sim N\left(\sum_{j\in pa_i} w_{ij}X_j + b_i,\ v_i\right) \tag{1}$$

where w_{ij} and b_i are parameters that govern the relationship of $X_{pa_i} \to X_i$, and v_i is the conditional variance of X_i.

A DAG 'G' depicts a set of variables, $X_1,X_2,\ldots,X_m$, and their relationships. Each node of G is a variable, and the directed arcs represent the parental relationships among the variables. The joint probability distribution of G is

$$p(x_1,x_2,\ldots,x_m|G) = \prod_{i=1}^{m} p(x_i|x_{pa_i},G) \tag{2}$$

where $p(x_1,x_2,\ldots,x_m|G)$ represents the probability distribution of a specific combination $x_1,x_2,\ldots,x_m$ from the variables $X_1,X_2,\ldots,X_m$, and x_{pa_i} is a vector that represents the list of direct parents of X_i, as depicted by G. BNs are locally structured, meaning that each node interacts only with its parent nodes.

For a given structure G under consideration, we estimate its parameters in Eq. (1) and derive the distribution of the latent variable Z_j. Because our graphical model includes a latent variable, the existing method cannot be applied. So, we developed a new procedure, which will be described in Section 3.

2.2. Learning Bayesian networks from data

If the structure of the BN is not known, the underlying structure of the BN given by G must be learned. This structure includes the specifications for the conditional independence assumptions among the variables of the model and the parameters. Many DAGs may determine the same joint probability distribution. Therefore, the family of all DAGs with a given set of vertices is naturally partitioned into Markov-equivalence class, each class being associated with a unique statistical model. This means that the structure of

the network can only be learned up to its Markov equivalent class. The literature describes two approaches applied to the task of learning the network structure from data: a constraint-based method and a score-based method. The recent constraint-based method is divided into two steps: the Markov Blanket (MB) discovery step and the causality discovery step. The MB discovery algorithms include Total Conditioning (TC) [27], IAMB [33], IPC-MB [14], and DOS [36]. The causality discovery algorithms include GS [24], CS [28], MS [7], and AO [4].

Here, we use TC for MB discovery which has the lowest time complexity and the revised version of MS which will be described in Section 3. The constraint-based methods [26,32] are efficient, but they may not be robust in terms of accuracy and may encounter difficulties if no causally sufficient condition exists. An algorithm's outcome is extremely sensitive to failures of conditional independency statistical tests. Furthermore, the conditional independency statistical tests cannot be easily used if any two or more variables have a hidden cause [32]. The score-based methods can remedy the demerits of constraint-based methods, although they are relatively inefficient computationally. Thus, the approach proposed in this paper adopts mainly a constraint-based method but supplemented by a score-based method if the causal direction is not determined. Several scoring functions can be considered, including the Bayesian scores based on the posterior probability of G [8], approximations of the posterior probability based on the Bayesian Information Criteria (BIC) [31], or an information-geometric score that considers the volume of the manifold that maps the underlying statistical model [29,2,19].

Once the structure of the BN becomes known, the maximum likelihood estimator (MLE), maximum *a posteriori* (MAP), or a Bayesian estimator can be used to estimate the graph parameters [17,5,6]. Here, we use the regression analysis, which is similar to MLE when the sample size is large enough, assuming that the latent variable has a Gaussian prior.

3. Proposed method

The proposed procedure of ranking evaluation consists of two phases which are repeated until the network structure converges. First, rankings are estimated using Gibbs sampling and gradient descent from a given network structure. Second, an updated network structure is found by the revised Multiple Search (MS) algorithm using the given rankings. The revised MS algorithm will be described in Section 3.2 and Appendix A. The final rankings are obtained based on the final graph structure.

The flow chart of the proposed method is shown in Fig. 2. The initial structure G is obtained from the function InitialStructure(X_i) which is the same as the function FindStructure(Z_j, X_i) except for using observable variables X_i's only. From the function GetRankings(G), we estimate a 'universal' latent variable from the given structure G because we assume that a latent variable affects all observable variables. Then, from the function FindStructure(Z_j, X_i), we obtain the structure G for the data and the estimated latent variable. Finally, these two functions are iterated until the termination conditions are met. The functions GetRankings(G) and FindStructure(Z_j, X_i) will be described in Sections 3.1 and 3.2, respectively. The iterative procedure is similar to the usual EM algorithm [9], but the difference is that FindStructure(Z_j, X_i) may not generate a structure having the maximum score because of the constraint-based algorithm used. If the proposed procedure does not converge within the specified number of iterations, the final rankings are drawn for the structure having the highest score so far. The pseudo-code of the proposed procedure is shown in Fig. A2.

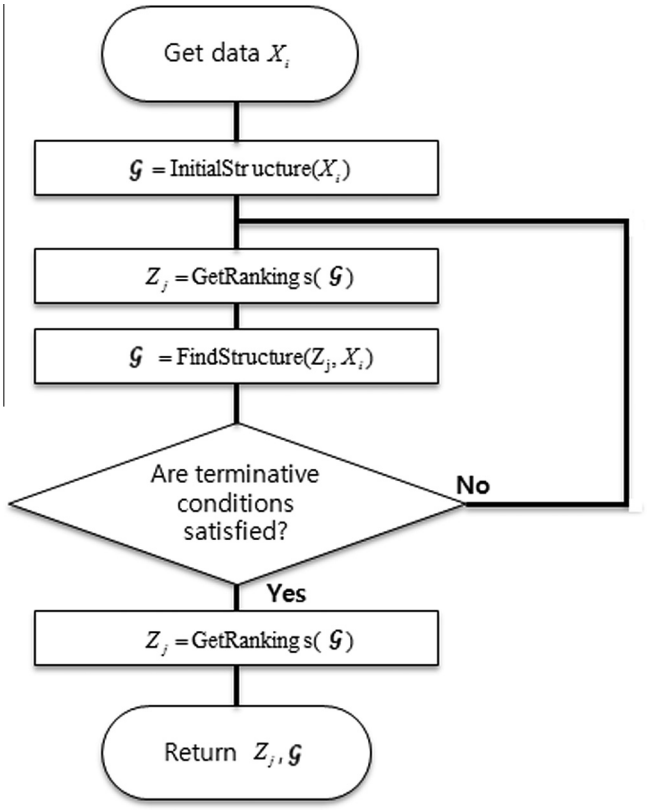

Fig. 2. Procedure of the proposed method.

3.1. Ranking estimation through GetRankings(G)

Suppose that variable X_i ($i = 1,2,\ldots,m$) has a vector of k_i parents $x_{pa_i} = \left(X^1_{pa_i}, X^2_{pa_i}, \ldots, X^{k_i}_{pa_i}\right)$. Also, assume the following linear Gaussian model for X_i given its parents and Z_j:

$$X_i|(x_{pa_i}, Z_j) \sim N\left(\beta^i_0 + \beta^i_z Z_j + \beta^i_1 X^1_{pa_1} + \cdots + \beta^i_{k_i} X^{k_i}_{pa_1}, \sigma^2_i\right) \quad (3)$$

where β^i_z is a coefficient that describes the relationship between X_i and Z_j, $\beta^i_t (t = 1,\ldots,k_i)$ is a coefficient that describes the relationship between X_i and $X^t_{pa_i}$, and σ^2_i is the conditional variance of X_i. In fact, we are considering the regression model for each variable. Once the value of Z_j is known, the regression model can be estimated easily. However, Z_j is not known, so we must estimate it and the regression coefficients simultaneously. We propose an iterative method through the function **GetRankings**(G) for estimating values of the latent variable and the regression parameters simultaneously by assuming *a priori* that each latent variable follows a standard normal distribution.

GetRankings(G) consists of six steps as in Fig. 3, and it is repeated until enough samples for Z_j are attained. **Step 1**; generate the prior of Z_js from a standard normal distribution. **Step 2**; from Z_j, the regression model $\left(\beta^i_0, \beta^i_z, \beta^i_t (t = 1,\ldots,k_i), \sigma^2_i\right)$ and is estimated easily using the ordinary least square method (OLS). **Step 3**; obtain a new Z_j to maximize the posterior distribution of Z_j because the complete joint posterior distribution of all parameters is complex. **Step 4**; repeat steps 2 and 3 to calculate the mean value of the produced Z_js. **Step 5**; distribution of Z_j is obtained by repeating steps 1–4, where one sample of Z_j comes from one iteration. **Step 6**; the median of the posterior distribution of Z_j is used as the final estimate of Z_j, and the other parameters are obtained through step 2 based on the final estimate of Z_j.

Function **GetRankings**(
G /*Bayesian network structure*/
Data /*Observable variable data; it is dropped in function expression*/
)
for k = 1 to 1000 do
generate Z_j s(0) from N(0, 1) /*step 1*/
for g = 1 to 100 do
Estimate β^i and σ_i for each X_i using Z_j s and regression analysis based on *G*; /*step 2*/
Estimate Z_j^{new} s using Eq. (7); /*step 3*/
Z_j s(g) = Standardized Z_j^{new} s;
endfor;
Z (k) = mean(Z_j s (51:100)); /*step 4*/
endfor; /*step 5*/
Estimate β^i and σ_i for each using median of Z (k) and regression analysis based on *G*; /*step 6*/
Return median of Z (k);

Fig. 3. Function **GetRankings**(*G*).

Now, we provide the theory behind Step 3, which is the key step. The conditional probability density function (pdf) of Z_j given observed data and regression estimators can be expressed by

$$p(z_j|x_1,\ldots,x_m) \propto p(z_j)p(x_1,\ldots,x_m|z_j) \propto exp\left(-\frac{z_j^2}{2}\right) \times \exp\left(-\sum_{i=1}^{m}\frac{\left(x_i-\left(\beta_0^i+\beta_z^i z_j+\beta_1^i x_{pa_i}^1+\cdots+\beta_z^i x_{pa_i}^{k_i}\right)\right)^2}{2\sigma_i^2}\right) = \exp\left(-\frac{z_j^2}{2}-\sum_{i=1}^{m}\frac{\left(x_i-\left(\beta_0^i+\beta_z^i z_j+\beta_1^i x_{pa_i}^1+\cdots+\beta_{k_i}^i x_{pa_i}^{k_i}\right)\right)^2}{2\sigma_i^2}\right) \quad (4)$$

To get Z_j yielding the maximum conditional pdf, the following optimization problem can be considered:

$$\text{Minimize } J=\frac{z_j^2}{2}+\sum_{i=1}^{m}\frac{\left(x_i-\left(\beta_0^i+\beta_z^i z_j+\beta_1^i x_{pa_i}^1+\cdots+\beta_{k_i}^i x_{pa_i}^{k_i}\right)\right)^2}{2\sigma_i^2} \quad (5)$$

Then, according to a gradient descent method, value of Z_j is updated for given the other parameters by

$$z_j^{new}=z_j^{old}-\eta\frac{\partial J}{\partial z_j}=z_j^{old}-\eta\left(z_j^{old}-\sum_{i=1}^{m}\frac{\beta_z^i}{\sigma_i}\frac{(x_i-\mu_i)}{\sigma_i}\right)$$

where $\mu_i=\beta_0^i+\beta_z^i z_j^{old}+\beta_1^i x_{pa_i}^1+\cdots+\beta_{k_i}^i x_{pa_i}^{k_i}$. If we set learning rate η as 1, then

$$z_j^{new}=\sum_{i=1}^{m}\frac{\beta_z^i}{\sigma_i}\frac{(x_i-\mu_i)}{\sigma_i} \quad (6)$$

The gradient descent is used here rather than taking the posterior mean because the result tends to overfit about only one variable when the posterior mean is used. Finally, we estimate the new latent variable as

$$z_j^{new}=\sum_{i=1}^{m}\frac{\beta_z^i}{\sigma_i}\frac{\left(x_i-\left(\beta_0^i+\beta_z^i z_j^{old}+\beta_1^i x_{pa_i}^1+\cdots+\beta_{k_i}^i x_{pa_i}^{k_i}\right)\right)}{\sigma_i} \quad (7)$$

The form of this result is similar to that in Guarino et al. [16]. After iterations, the distribution of Z_j in step 5 will be used for the ranking of institutions. The term $\frac{\beta_z^i}{\sigma_i}$ in Eq. (7) can be used as an index of the relative importance of each variable, and it can specify variables that are more influential than others on the latent variable Z_j. That is,

$$\text{Relative importance of variable } i=\frac{\beta_z^i}{\sigma_i} \quad (8)$$

The location and scale of the Z_js are not identifiable because any shift in the Z_js can be offset by a shift in β_0^is and any rescaling of the Z_js can be offset by rescaling β_z^is. This means that the observed data need not be shifted or rescaled because these processes are built into the model.

The regression parameters and Z_js are estimated jointly. **GetRankings**(*G*) alternates between computing the conditional distribution of the Z_js and computing the conditional distribution of the regression parameters. This is basically the Gibbs sampling method. The latent variable rankings can be derived by using a Gibbs sampler to draw samples from the distribution of the Z_js given the observed data. Rankings can also be obtained by comparing estimates of the distribution of the Z_js (i.e., the ranking score for each institution), and an interval around Z_j (similar to a confidence interval) can be obtained by repeating the process 1000 times in step 5, using the 5th and 95th percentiles which cover a 90% posterior probability interval; in step 4, we take 100 samples, remove the first 50 samples and average the remaining 50 to remove warm-up and get one sample of the distribution of Z_j. The number of samples can be changed depending on the data.

3.2. Structure learning through FindStructure(Z_j, X_i)

For given latent variable rankings Z_js, we learn the network structure G related to Z_j and X_i. We use the TC algorithm [27] as MB discovery and the revised MS as causality discovery. Although these algorithms cannot be applied when there are latent variables, we apply them after estimating the latent variable Z_j by **GetRankings**(G).

The revised MS corrects some hidden problems in the MS algorithm [7], which is expected to have better accuracy although it has higher time complexity. MS carries out the conditional independence test for finding V-structure by using the p-value of each independent variable in a linear regression model as a criterion. However, the p-value CI test may fail for loosely connected spouse nodes if a large noise exists between two nodes. Consequently, MS could cause the double-check error for V-structures, which could generate the spouse problem and the cycle problem. Definitions of these terms and how to revise this problem are explained in Appendix A.

The revised MS gives a partially directed acyclic graph (PDAG). Therefore, determining the direction of an arc, if undecided, is a major issue here. To determine the direction of an undirected edge, we first orient maximally a PDAG by the revised MS using 3 rules (Fig. A4). It is common in structure learning to obtain a completed PDAG [26] from a constraint-based algorithm. Meek [25] has proved that these three rules indeed return the maximally oriented graph when given a PDAG whose V-structures are oriented. However, there could still be undirected edges. Here, we use the score evaluation to decide the direction, which is called **GetScore**(G). For given graph G, **GetScore**(G) returns the BIC score using the parameters and the latent variable Z_j from **GetRankings**(G). That is, the function **GetScore**(G) evaluates the structure G using the score BIC that is based on the network's joint posterior probability distribution. The following score is calculated for given structure G to be considered:

$$\textbf{GetScore}(\boldsymbol{G}) = -1 * \text{BIC score} = -1 * (-2\ln L + s\ln(n)) \quad (9)$$

where L is the maximized value of the likelihood function for the estimated model, and s is the number of free parameters to be estimated.

In summary, **FindStructure**(Z_j, X_i) consists of four steps and returns a completed directed acyclic graph (DAG) as the network structure G (Fig. 4 and Fig. A3). **Step 1**; find MB for each node using TC feature based on data including Z_j and X_i. Here, we assume that Z_j has all X_i as MB. **Step 2**; determine the direction of edges in a V-structure in the network using the revised MS and MB information of step 1. **Step 3**; orient a partially directed acyclic graph maximally using 3 rules. **Step 4**; if there is an undirected edge after step 3, choose the direction having a higher score by comparing two possible directions for each edge through the function **GetScore**(G). Throughout this paper we used $\alpha = 0.05$ as the significance level for independency and $\beta = 0.01$ as the significance level for dependency.

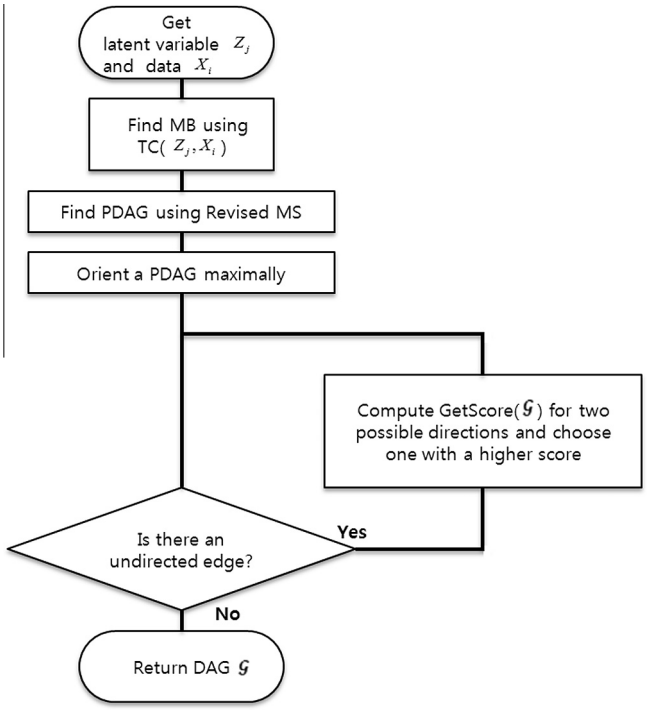

Fig. 4. Flow chart of **FindStructure**(Z_j, X_i).

3.3. Example

To illustrate how the proposed procedure (Fig. 2) works in evaluating the rankings, we consider the artificial data for 10 institutes having seven indicators (Var A to Var G) as shown in Table 1. This data was generated based on Network 1 in Fig. 10 to which one directed edge is added from Var F to Var G.

The initial graph is obtained from the function InitialStucture(X_i), which follows Steps 1–3 of the function FindStructure(Z_j, X_i) using only the data of X_i's. The resulting graph is shown in Fig. 5. For the initial graph the direction of each undirected edges is chosen arbitrarily.

Now, the iterative procedure is applied to improve the graph having a higher score until the termination condition is met. At each iteration, the latent variable is computed for each institute, and the rankings are made by sorting the values of the latent variable. Table 2 shows the result of the first iteration and Fig. 6 shows the resultant graph of which calculated score is 18.08.

For this example, the final result is shown in Table 3 and Fig. 7. The final score obtained is 175.16.

Table 1
Ten institutes data for our example.

Institute	Var A	Var B	Var C	Var D	Var E	Var F	Var G
1	0.167616	−0.02301	−34.7216	−5.41621	3.419723	−0.48212	−4.2994
2	2.458481	6.77641	296.3316	48.91263	9.683911	6.548031	64.84344
3	−5.40795	−16.1454	−739.515	−121.146	−24.5906	−15.5332	−154.327
4	0.53123	1.74581	76.28637	12.56782	2.803942	1.643768	16.049
5	−0.81186	−2.22682	−52.6482	−8.80762	−10.5758	−1.31626	−12.9375
6	−3.2786	−10.6913	−532.888	−86.8208	−11.3265	−10.8149	−107.624
7	−2.3118	−5.91282	−251.115	−41.7309	−8.45308	−5.7478	−57.1094
8	−0.85694	−1.45301	−68.2647	−11.8455	2.394036	−1.9208	−18.6746
9	5.18679	16.39727	771.1489	125.6851	25.75457	15.76694	157.208
10	3.021252	11.56167	606.4878	98.12875	12.51212	11.70282	116.7579

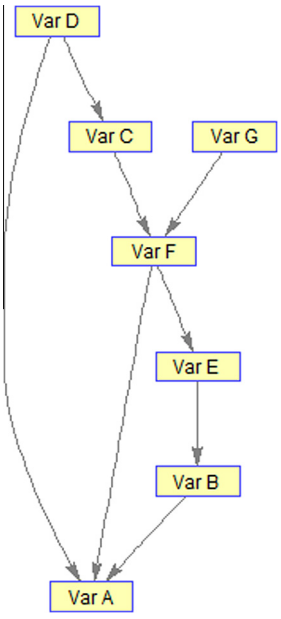

Fig. 5. Initial graph for the example.

Table 2
First iteration for the example.

Institute	Latent value	Ranking
1	−1.78459	10
2	1.36388	1
3	−0.25991	7
4	0.121127	5
5	−1.37478	9
6	0.031892	6
7	−0.3601	8
8	1.056166	2
9	0.535894	4
10	0.67042	3

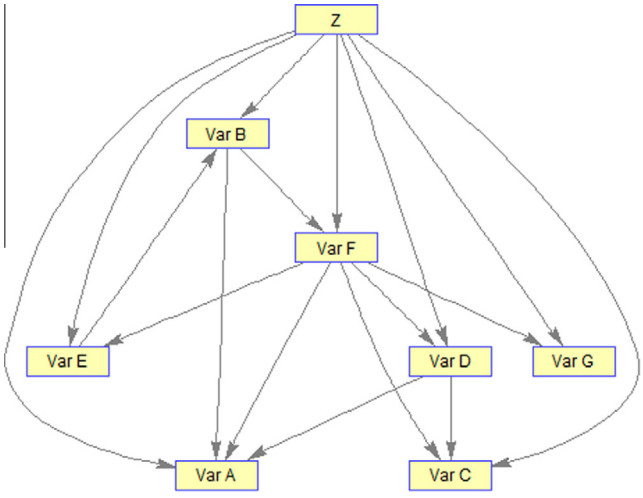

Fig. 6. The graph from the first iteration for the example.

4. Application

The real data for this application are those used by the Joongang Daily Newspaper to evaluate rankings for 41 Industrial Engineering (IE) departments in Korea. Thirteen indicators (variables) were considered and they are abbreviated (SFR, CR, DJ, IJ, KJ, SJ, BS, ORF, SCH, IRF, GE, DR, RR) as shown in Table 4. Before the analysis, each variable was standardized to have mean 0 and variance 1. In addition, the indicators SFR and DR were multiplied by −1 so that all variables have the property that higher means better.

Table 3
Final result of the example.

Institute	Latent value	Ranking
1	−0.12042	5
2	0.585237	3
3	−1.58084	10
4	0.144007	4
5	−0.13776	6
6	−1.17299	9
7	−0.53302	8
8	−0.14266	7
9	1.633665	1
10	1.324769	2

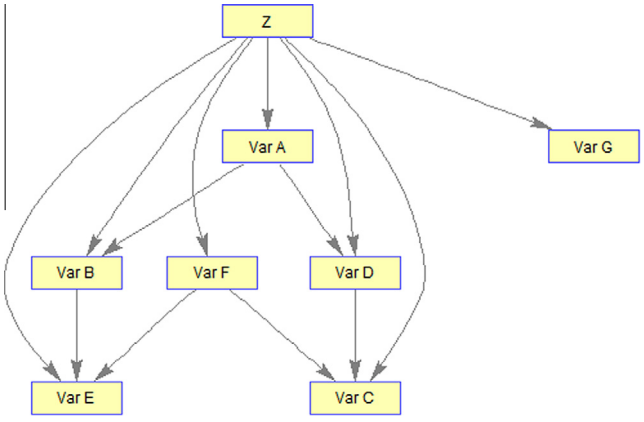

Fig. 7. Final graph of the example.

We apply the proposed method to the data to obtain a network of variables and the ranking of 41 departments. The maximum number of iterations in the proposed method was limited to 40 when the algorithm fails to converge. The network derived using the proposed method (Fig. 8) shows some significant directed arcs indicating causality among variables. The number of SCI journal publications (SJ) and the number of international journal publications (IJ) have numbers of outgoing arrows larger than others, which means that most of the variables are affected by SJ and IJ in the graph.

The estimated rankings of 41 IE departments in Korea were presented as intervals as well as medians, and they were ordered by median rankings so that top-ranked departments appear first (Fig. 9). The proposed method evaluates the ranking of each institution 1000 times before the ranking distribution of each institution is obtained. The median ranking is the median of this distribution, and the interval is between the 5th and 95th percentiles of the distribution. The intervals for top-ranked departments were narrower than those of departments near the middle and bottom.

Fig. 9 also includes the result by the sum method based on the sum of all indicators, which is a naïve estimator of ranking. The rankings by the proposed method and those by the sum were quite different. About 50% of the rankings by the sum were outside the intervals based on the proposed method. As will be seen in Section 5, the sum method works quite well when all the indicators have positive correlations with each other, but it may not work well when some variables have negative correlations as in this case.

Table 4
Indicators used for ranking IE departments.

Evaluation category	Indicators	Abbreviated form	Equation
Educational environment	Student to fulltime faculty ratio	SFR	Number of student/number of faculty full-time equivalents
	Percentage of classes by fulltime faculty	CR	(Classes by faculty/number of total classes) * 100
Research capability	Number of domestic journal publication per faculty	DJ	University official data
	Number of international journal publication per faculty	IJ	University official data
	Number of KRF journal publication per faculty	KJ	University official data
	Number of SCI journal publication per faculty	SJ	University official data
	Number of books written per faculty	BS	University official data
	Outside research funds per faculty	ORF	Funds from governments, private organization and foreign institutions
Funding	Scholarship per student	SCH	Average scholarship for a semester
	University inside research funds per faculty	IRF	Research grants from its own university
Education effect	Percentage of graduates employed	GE	University official data
	Dropout ratio	DR	(Number of dropout/number of students on register) * 100
	Freshman recruitment rate	RR	(Number of incoming students/number authorized) * 100

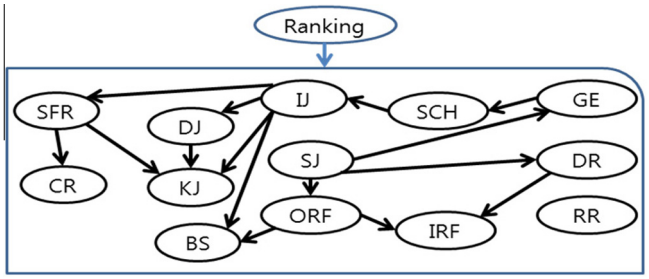

Fig. 8. Derived network for evaluating IE departments.

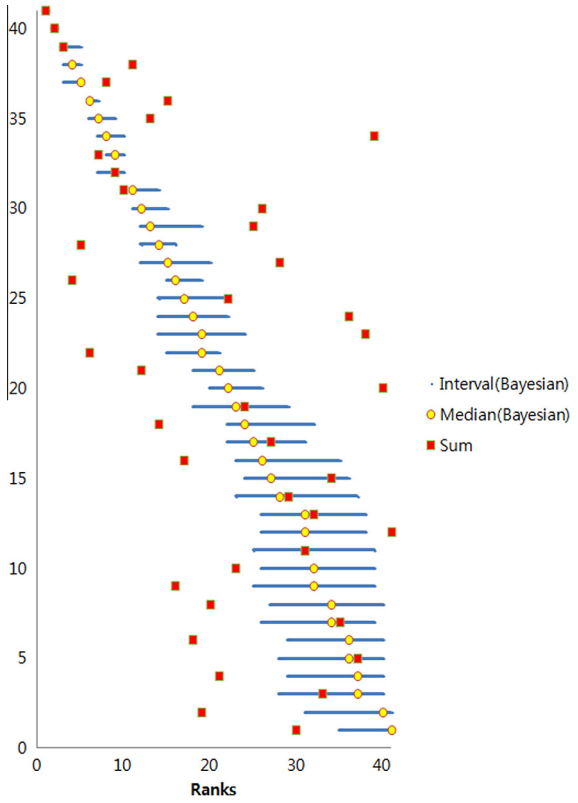

Fig. 9. Estimated rankings for IE departments using real data.

Table 5
The relative importance of each indicator.

Indicator	Mean of Importance	Variance of Importance
SFR	0.126	1.00×10^{-4}
CR	0.031	9.25×10^{-6}
DJ	0.353	5.42×10^{-4}
IJ	**0.766**	**2.18×10^{-4}**
KJ	0.295	4.18×10^{-4}
SJ	**0.976**	**6.69×10^{-4}**
BS	0.039	7.71×10^{-6}
ORF	0.055	3.84×10^{-5}
SCH	0.244	7.92×10^{-6}
IRF	0.077	1.24×10^{-5}
GE	0.142	3.36×10^{-4}
DR	0.133	2.19×10^{-4}
RR	0.004	1.43×10^{-9}

Note: Bold numbers show the two most important indicators.

As defined in Eq. (8), the relative importance of each variable can be compared. The mean and variance of the relative importance weight of each indicator were calculated for this real data (Table 5). The number of SCI journal publications (SJ) and the number of international journal publication per faculty (IJ) seem to have greater influence on the ranking than do other variables. This result is consistent with the graph shown in Fig. 8. Using a Bayesian network allows an institution to realize which variables are important to rankings given its current situation and to perform what-if analyses for improving the ranking. However, we have not performed this analysis here because it is beyond the scope of this paper.

5. Experiment for validation

We performed two experiments on artificial data to check the accuracy and consistency of the proposed method. The two true networks (Networks 1 and 2) are given in Fig. 10. Network 1 is composed of seven variables (variables $A,B,\ldots,G$) and Network 2 has 15 variables. Both networks have one latent variable Z. The coefficients used when generating variables were randomly chosen from the uniform distribution between 2 and 10. For example, we first generated ranking scores Z_j of 50 institutions randomly from a standard normal distribution, and the actual rankings of these institutions were obtained by comparing these scores. Then, we generated 50 observations on each variable in Network 1, for example, from the following distributions:

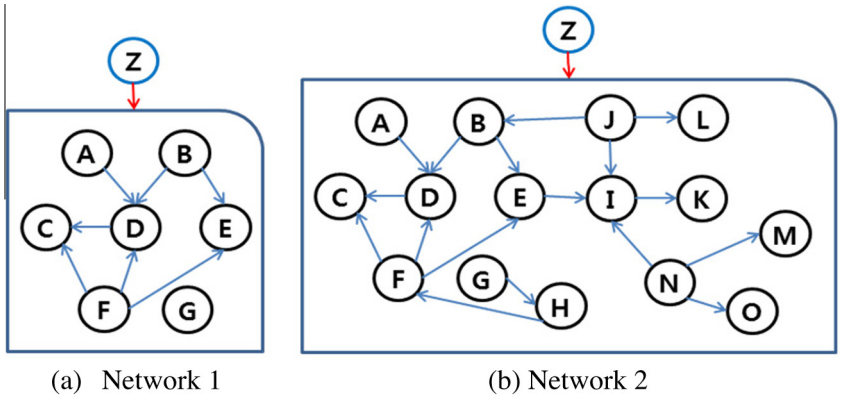

Fig. 10. True network structure assumed.

Table 6
Experiment result for artificial data.

	σ^2	Proposed method		MS	
		# Total error	Time	# Total error	Time
Network 1	0.1	4 ± 0 (57%)	348.3 ± 33.0	7.66 ± 1.53 (109%)	0.018 ± 0.000
	0.5	2.8 ± 1.30 (40%)	435.3 ± 201.4	5.8 ± 2.17 (83%)	0.033 ± 0.002
Network 2	0.1	9.25 ± 0.96 (54%)	491.8 ± 286.6	12 ± 0.82 (71%)	0.032 ± 0.002
	0.5	9.5 ± 0.71 (56%)	436.0 ± 410.1	12 ± 1.41 (71%)	0.062 ± 0.003

Note: The figures in parenthesis are percentages relative to the total number of edges.

Variable $A \sim N(3.9Z, \sigma^2)$

Variable $B \sim N(4.2Z, \sigma^2)$

Variable $C \sim N(4.3Z + 3.5D + 2.9F, \sigma^2)$

Variable $D \sim N(4.6Z + 4.5A + 3.3B + 3.8F, \sigma^2)$

Variable $E \sim N(2.5Z + 3.2B + 4.2F, \sigma^2)$

Variable $F \sim N(2.6Z, \sigma^2)$

Variable $G \sim N(3.7Z, \sigma^2)$

All measurable variables had the same variance σ^2. Two cases of σ^2 were considered ($\sigma^2 = 0.1$ and 0.5). The second experiment considered fifteen variables, and the remainders were the same. The maximum number of iterations in the proposed method was limited to 20 in the case of no convergence. We repeated the experiment 5 times.

We used the structure error measure as in Choi and Jun [7] to compare the structures derived by the proposed method using only the observable variables based on the true structures. The results are summarized in Table 6, where the total structure error includes four types: edge addition, deletion, reversal, and direction missing. It was found that MS does not identify the causal relation well when an unobservable latent variable is present. However, the proposed method outperformed MS in terms of structural accuracy although time complexity is by far larger than MS.

The 50 institutions were ranked in the form of 90% confidence intervals, and they are reported in Fig. 11 ($\sigma^2 = 0.1$) and Fig. 12 ($\sigma^2 = 0.5$) for Network 1, while in Fig. 13 ($\sigma^2 = 0.1$) and Fig. 14 ($\sigma^2 = 0.5$) for Network 2. Most of the actual rankings fell within the 90% confidence intervals. Furthermore, the interval lengths were narrow enough to distinguish some institutions from other institutions statistically. These two findings demonstrate that the rankings of the top and bottom institutions are more certain than those in the middle. First, the middle-ranked institutions have larger intervals than the top or bottom-ranked institutions. Second, increasing the variance and the number of variables affected the result of the middle-ranked institutions, especially the interval length, more than the top or bottom-ranked institutions.

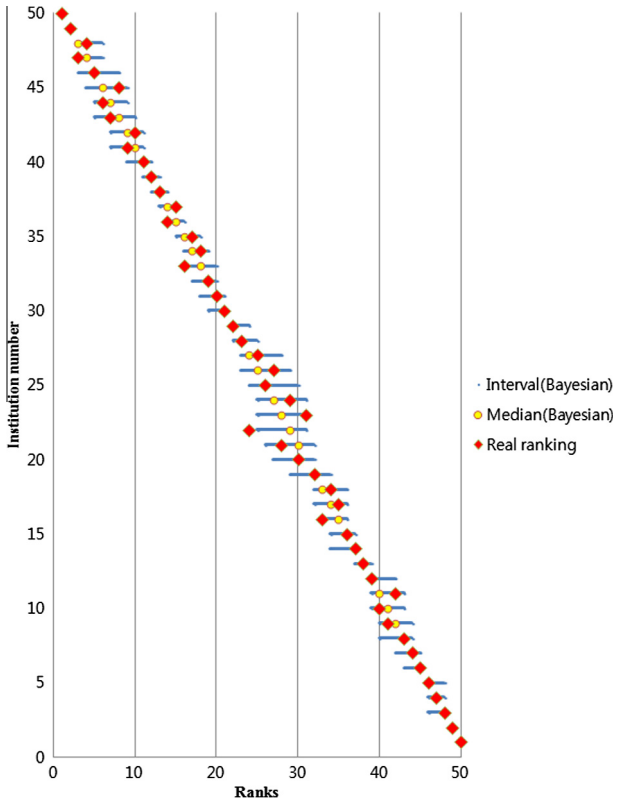

Fig. 11. Ranking estimation using artificial data (Network 1, $\sigma^2 = 0.1$).

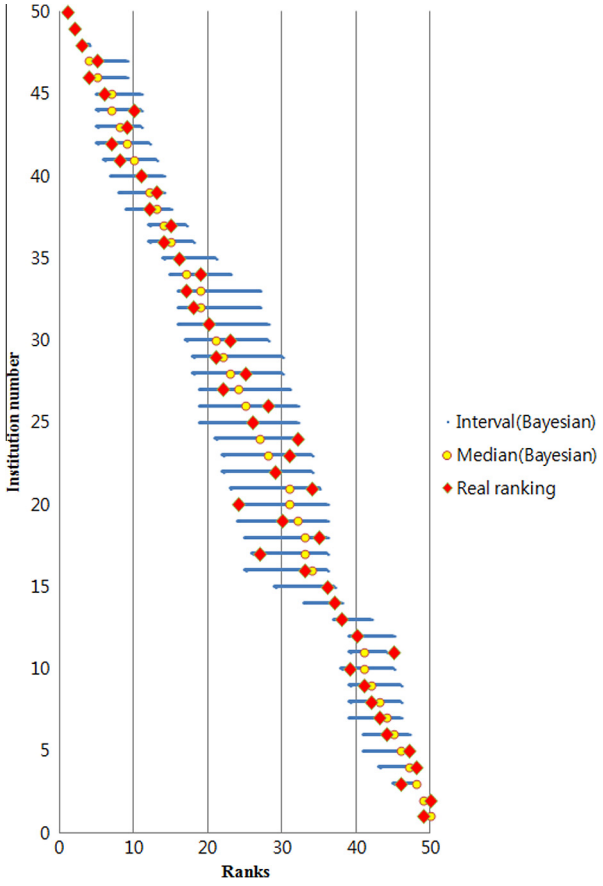

Fig. 12. Ranking estimation using artificial data (Network 1, $\sigma^2 = 0.5$).

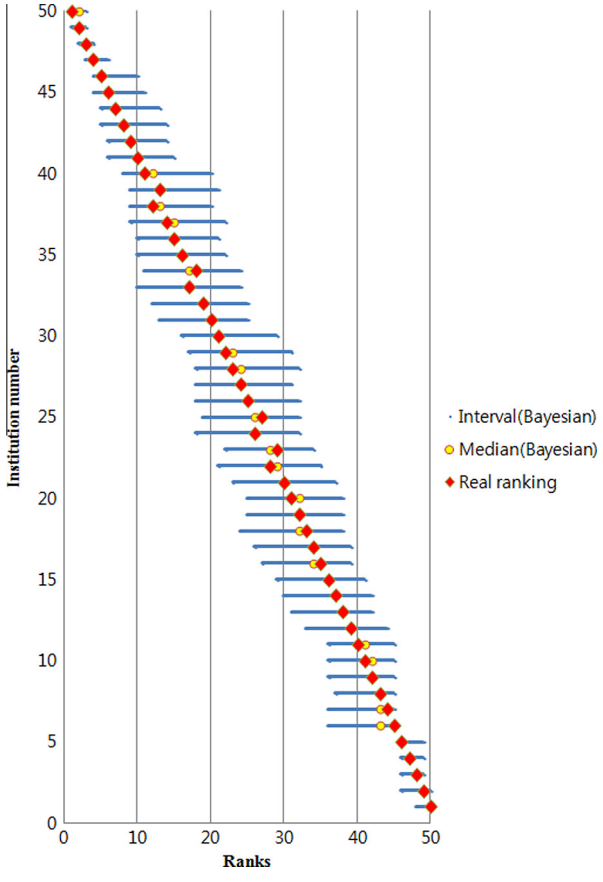

Fig. 13. Ranking estimation using artificial data (Network 2, $\sigma^2 = 0.1$).

Table 7
The comparison of ranking estimation for artificial data.

σ^2	Criterion	Proposed method	Sum method
Network 1 (7 variables)			
0.1	MAE	1.81 ± 1.60	1.95 ± 1.63
	Top 5 inclusion ratio	0.87 ± 0.12	0.93 ± 0.12
	Top 10 inclusion ratio	0.93 ± 0.12	0.93 ± 0.12
	Top 20 inclusion ratio	0.95 ± 0.09	0.95 ± 0.05
0.5	MAE	1.92 ± 0.50	2.36 ± 1.61
	Top 5 inclusion ratio	0.88 ± 0.18	0.88 ± 0.18
	Top 10 inclusion ratio	0.96 ± 0.05	0.88 ± 0.11
	Top 20 inclusion ratio	0.95 ± 0.05	0.93 ± 0.08
Network 2 (15 variables)			
0.1	MAE	0.74 ± 0.62	0.81 ± 0.46
	Top 5 inclusion ratio	0.95 ± 0.10	0.90 ± 0.12
	Top 10 inclusion ratio	1.00 ± 0.00	1.00 ± 0.00
	Top 20 inclusion ratio	0.96 ± 0.08	0.96 ± 0.05
0.5	MAE	0.91 ± 0.64	1.12 ± 0.17
	Top 5 inclusion ratio	1.00 ± 0.00	1.00 ± 0.00
	Top 10 inclusion ratio	0.95 ± 0.07	1.00 ± 0.05
	Top 20 inclusion ratio	0.98 ± 0.03	0.95 ± 0.05

Table 8
The result of ranking estimation with smaller observations.

Measures	Proposed method	Sum method
MAE	1.11 ± 0.73	1.88 ± 2.53
Top 5 inclusion ratio	0.8 ± 0.35	0.73 ± 0.46

The median rankings differ from the actual rankings; more often σ^2 is larger. However, increasing the number of variables reduces the difference between the median rankings and the actual rankings. The median rankings and the actual rankings can be compared based on some measures. We first considered the following mean absolute error (MAE) of rankings:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|r_i^{true} - \hat{r}_i\right| \tag{10}$$

where r_i^{true} is the actual ranking and $\hat{r}_i$ is the estimated median ranking. Second, we considered the Top-K inclusion ratio, which is the proportion of the institutions in the estimated Top-K rankings that are in the actual Top-K ranking. We considered $K = 5$, $K = 10$, and $K = 20$. We compared the result of the proposed method with that of the sum method, which simply sums up all variables and uses this sum to estimate the ranking. These comparisons are summarized in Table 7. Generally, the sum method gives a good result in case of all positive coefficients. However, the sum method is expected to perform worse if some coefficients have negative signs. It is shown that the proposed method estimates ranks more accurately than the sum method. When σ^2 increased, the estimation performance by the sum method degraded slightly. However, the results of the proposed method were relatively stable. It is also observed that increasing the number of variables has a good effect on the estimation performance.

Lastly, we checked how reliable the proposed method is when the number of observations is small. We set up the experiment with Network 1 ($\sigma^2 = 0.1$), but now generating variables for only 10 institutions as observations. It was found that the proposed method yields the total number of error of 3.33 ± 0.58 as compared with 6.67 ± 1.15 when using MS. Table 8 and Fig. 15 show the result, which shows that the proposed method still works with smaller observations.

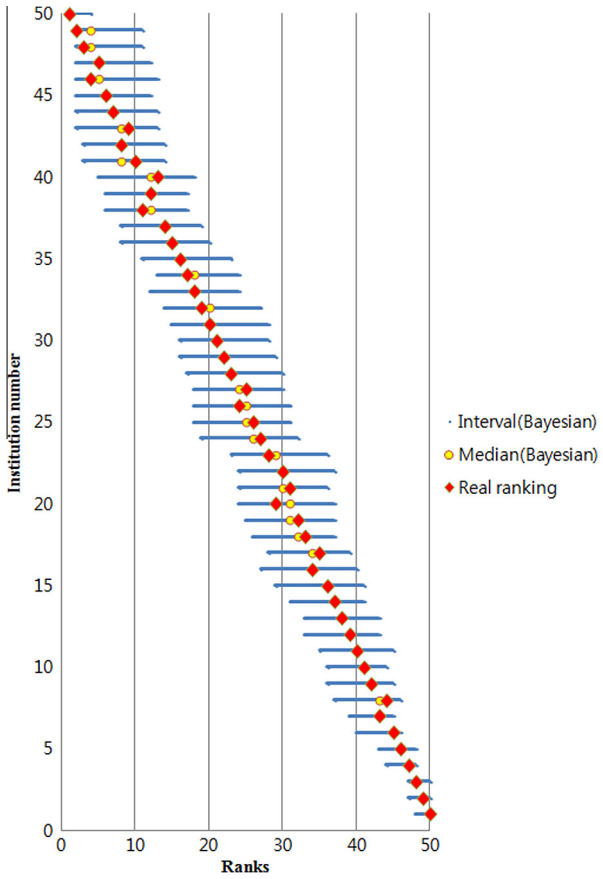

Fig. 14. Ranking estimation using artificial data (Network 2, $\sigma^2 = 0.5$).

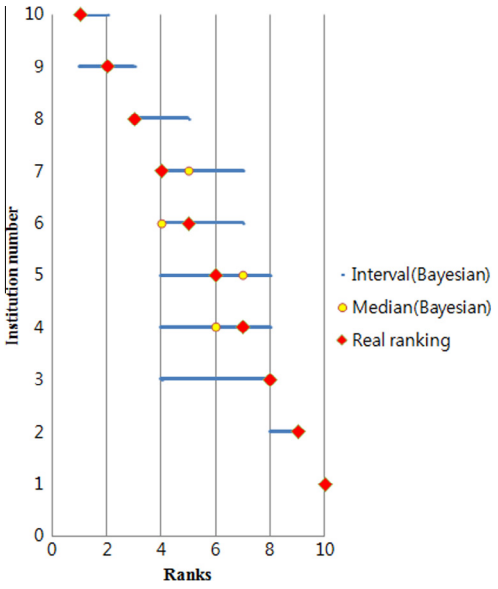

Fig. 15. Ranking estimation with smaller observations (Network 1, $\sigma^2 = 0.1$).

6. Conclusion

In this paper, we considered a new type of Bayesian network containing an unobservable latent variable which affects all observable variables. We applied the proposed method to the task of ranking institutions and determining the rankings in terms of intervals based only on the institutions' scores in the performance indicators. The linear Gaussian models are utilized to express the causal relationship among variables in the network. The proposed iterative method finds the network structure of variables by using a combined algorithm of the revised MS and the BIC scoring function, estimating the parameters by using Gibbs sampling, gradient descent and regression analysis.

The latent variable that represents the ranking score of an institution is estimated, and the final rankings are determined by comparing those scores. The interval estimate and the median estimate of the ranking of an instruction are obtained by repeating the estimation process. The analyses of a real case and some artificial data sets demonstrate the performance of the proposed method.

To sum up, the proposed method has some advantages as a ranking system. First, the proposed method compresses the information of data (performance indicators in our case) into a single latent variable. Second, the proposed method does not require weights for indicators in advance, and therefore, it avoids the controversial subjective assignment of such weights. Third, using the interval estimate rather than the point estimate makes it possible to compare institutions statistically.

However, the proposed method still has some limitations in computation time and in convergence. We also need to handle some semantic constraints in the network structure when utilizing an expert knowledge. On the other hand, the proposed method can be applied to a more general model where a latent variable has both in-degree and out-degree to all observable variables. If we consider a latent variable as a clustering node or a class node, then we can apply the proposed method to an unsupervised clustering problem or a classification problem. These issues should be considered in the future work.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea from the Ministry of Education, Science and Technology (Project No. 2012-0001665).

Appendix A. The revised MS algorithm

The MS algorithm [7] is known to be fast and accurate but it can cause double-check errors for V-structures. The double-check error means that for a V-structure $X \rightarrow Y \leftarrow Z$ can be detected from the CI test using X as the dependent variable while it cannot be detected from the CI test using Z as the dependent variable. The double-check error may generate a spouse problem and a cycle problem. The spouse problem occurs when a spouse link of a V-structure $X \rightarrow Y \leftarrow Z$ coexists in another V-structure like $M \rightarrow X \leftarrow Z$. The cycle problem means that V-structure edges make a cycle. We cannot remove a spouse link from step 7 of phase 1 in the MS algorithm under these problems.

The revised MS treats these problems by adding steps to the phase 2 of MS (Fig. A1). For $X \rightarrow Y \leftarrow Z$, we test the effect of center Y using conditioning set $D(X, Y) \cup D(Y, X)$, where **D** is a conditioning set that d-separates X from Y, rather than $D(X, Y)$ because the former increases the probability of d-separation from X to Y. In the case of a spouse problem, we check again the characteristics of a V-structure and compare the degree of independency like the opposite direction collision case which is given in MS. Consequently, one of the two is deleted in the graph. In the case of a cycle problem, we test and compare every V-structure related to each edge in a cycle. As a result, the maximum degrees of

/* opposite direction collision; it is identical with MS */

0. **K**, **Vstr**, **D**() are given from Phase 1. /* Notation : **K**:skeleton of structure, **Vstr**:V-structure, **D(X,Y)**:a conditioning set that d-separates X from Y */

1. For two conflicting V-structures X→Z←Y and Z→X←W, if any, in **Vstr**, do Steps 2-5.

2. Set $S_1 = (D(X,Y) \cup D(Y,X)) \setminus \{Z\}$ and $S_2 = (D(Z,W) \cup D(W,Z)) \setminus \{X\}$.

3. If X and Y are conditionally dependent given S_1 or conditionally independent given $S_1 \cup \{Z\}$, remove X→Z←Y from **Vstr.**

4. If Z and W are conditionally dependent given S_2 or conditionally independent given $S_2 \cup \{X\}$, remove Z→X←W from **Vstr.**

5. If the degree of conditional independency of (X,Y) given S_1 is greater than that of (Z,W) given S_2, then remove Z→X←W from **Vstr**. Else, remove X→Z←Y from **Vstr.**

/* spouse problem */

6. For two conflicting V-structures X→Z←Y and W→X←Y, if any, in **Vstr**, do Steps 7-10.

7. Set $S_1 = (D(X,Y) \cup D(Y,X)) \setminus \{Z\}$ and $S_2 = (D(W,Y) \cup D(Y,W)) \setminus \{X\}$.

8. If X and Y are conditionally dependent given S_1 or conditionally independent given $S_1 \cup \{Z\}$, remove X→Z←Y from **Vstr.**

9. If W and Y are conditionally dependent given S_2 or conditionally independent given $S_2 \cup \{X\}$, remove W→X←Y from **Vstr.**

10. If the degree of conditional independency of (X,Y) given S_1 is greater than that of (W,Y) given S_2, then remove W→X←Y from **Vstr**. Else, remove X→Z←Y from **Vstr.**

/* cycle problem */

11. For a cycle $X_1, X_2, ..., X_n, X_{n+1}(= X_1); X_i \rightarrow X_{i+1}$ for i=1,…,n, if any, in **Vstr**, do Step 12-13.

12. For every V-structures $X_i \rightarrow X_{i+1} \rightarrow \mathrm{W}_j$ for j=1,…, m_i including $X_i \rightarrow X_{i+1}$ in **Vstr**, get the degree of conditional independency CI_{ij} of (X_i, W_j) given $S_{ij} = (D(X_i, W_j) \cup D(W_j, X_i)) \setminus \{X_{i+1}\}$ and maximum degree $MaxCI_i = \max_j CI_{ij}$.

13. If k = $\arg\min_i MaxCI_i$, then remove all $X_k \rightarrow X_{k+1} \rightarrow \mathrm{W}_j$ for j=1,…, m_k

14. Orient all edges in **Vstr** to **K** and remove spouse link in **Vstr** if any in **K.**

15. Return **K**

Fig. A1. Phase 2 of the revised MS algorithm.

```
Begin;
Obtain Data; /*Data: X_i information */
Initial G = InitialStructure( X_i )   /* Initial G is obtained from FindStructure only using */
while G is changed or terminative conditions are not satisfied do
        Z_j =GetRankings(G);
        G=FindStructure( Z_j, X_i );
end while
Z_j =GetRankings(G);
Return Z_j, G;
end;
```

Fig. A2. The pseudo codes of the proposed method.

independency *MaxCI* for each edge are obtained. Then, we remove all V-structure related to the edge which has the minimum value of *MaxCI*. It removes the edge which has a minimum effect of the center among the edges in a cycle. Lastly, we orient all the edges in V-structure to the structure and remove the spouse link in V-structure if there is a link because there are no more collisions or cycles.

Appendix B. Pseudo codes

Here we include pseudo codes of the proposed method (Fig. A2), FindStructure(Z_j, X_i) (Fig. A3), and OrientPDAGmax(G) (Fig. A4).

```
Function FindStructure(
Z_j     /*the value of Latent variable*/
X_i     /*data*/
)
Find MB using TC about data including Z_j and X_i ;     /*step 1*/
RevisedMS(MB);     /*step 2*/
OrientPDAGmax(G);     /*step 3*/
while there is an undirected edge do,     /*step 4*/
        compute two GetScore(G) about two possible directions;
        choose the direction which has a higher score;
end while
Return G
```

Fig. A3. FindStructure(Z_j, X_i).

References

[1] P. Aguilera, A. Fernandez, F. Reche, R. Rumi, Hybrid Bayesian network classifiers: application to species distribution models, Environmental Modelling and Software 25 (12) (2010) 1630–1639.
[2] S.I. Amari, Differential Geometrical Methods in Statistics, Springer-Verlag, Berlin, 1985.
[3] S. Bibi, I. Stamelos, Software process modeling with Bayesian belief networks, in: Proceedings of 10th International Software Metrics Symposium (Metrics 2004), Chicago, USA, 14–16 September 2004.
[4] A.T. Bui, C.-H. Jun, Learning Bayesian network structure using Markov blanket decomposition, Pattern Recognition Letters 33 (2012) 2134–2140.
[5] W. Buntine, Operations for learning with graphical models, Journal of Artificial Intelligence Research 2 (1994) 159–225.
[6] W. Buntine, A guide to the literature on learning probabilistic networks from data, IEEE Transactions on Knowledge and Data Engineering 8 (1996) 195–210.
[7] Y.-H. Choi, C.-H. Jun, A causal discovery algorithm using multiple regressions, Pattern Recognition Letters 31 (2010) 1924–1934.
[8] G. Cooper, E. Herskovits, A Bayesian method for the induction of probabilistic networks from data, Machine Learning 9 (4) (1992) 309–347.
[9] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society B39 (1) (1977) 1–38.
[10] N.E. Fenton, P. Krause, M. Neil, Software measurement: uncertainty and causal modelling, IEEE Software 10 (4) (2002) 116–122.
[11] N.E. Fenton, W. Marsh, M. Neil, P. Cates, S. Forey, M. Tailor, Making resource decisions for software projects, in: 26th International Conference on Software Engineering ICSE2004 May 2004, Edinburgh, United Kingdom, IEEE Computer Society, 2004, pp. 397–406.
[12] A. Fernandez, M. Morales, C. Rodriguez, A. Salmeron, A system for relevance analysis of performance indicators in higher education using Bayesian networks, Knowledge and Information Systems 27 (2011) 327–344.
[13] B.L. Fridley, E. Inversen, Y.Y. Tsai, G.D. Jenkins, E.L. Goode, T.A. Sellers, A latent model for prioritization of SNPs for functional studies, PLoS ONE 6 (6) (2011) e20764.
[14] S. Fu, M.C. Desmarais, Fast Markov blanket discovery algorithm via local learning within single pass, Proceedings of the Twenty-First Canadian Conference on Artificial Intelligence (2008) 96–107.
[15] H. Goldstein, D. Spiegelhalter, League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion), Journal of the Royal Statistical Society Series A 159 (part 3) (1996) 385–443.
[16] C. Guarino, G. Ridgeway, M. Chun, R. Buddin, Latent variable analysis: a new approach to university ranking, Higher Education in Europe 30 (2) (2005) 147–165.
[17] D. Heckerman, A tutorial on learning with Bayesian networks, in: M.I. Jordan (Ed.), Learning in Graphical Models, Kluwer Academic Publishers, Dordecht, The Netherlands, 1998, pp. 01–354.


```
Function OrientPDAGmax(
G     /*Bayesian network structure*/
)
while G is changed by some rule do
        for each X, Y, Z such that X→Y−Z do
                Orient as X→Y→Z /* no new V-structure */
        endfor
        for each X, Y such that X−Y and ∃ directed path from X to Y do
                Orient as X→Y     /* preserve acyclicity */
        endfor
        for each X, Y such that X−Y & ∃ nonadj. Z, W such that X−Z→Y & X−W→Y do
                Orient as X→Y     /* three fork V with married parents */
        endfor
end while
Return G
```

Fig. A4. OrientPDAGmax(G).

[18] N.M. Laird, T.A. Louis, Empirical Bayes ranking methods, Journal of Educational and Behavioral Statistics 14 (1) (1989) 29–46.
[19] E. Lauria, Learning the structure of a Bayesian network: an application of information geometry and the minimum description length principle, in: K.H. Knuth, A.E. Abbas, R.D. Morris, Patrick J. Castle (Eds.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering (2005) 293–301.
[20] E. Lauria, J. Duchessi, A methodology for developing Bayesian networks: an application to information technology (IT) implementation, European Journal of Operational Research 179 (1) (2007) 234–252.
[21] S.L. Lauritzen, Causal inference from graphical models, in: O.E. Barndorff-Nielsen, D.R. Cox, C. Kluppelberg (Eds.), Complex Stochastic Systems, CRC Press, London, 2000, pp. 3–107.
[22] K.F.R. Liu, C.F. Lu, C.W. Chen, Y.S. Shen, Applying Bayesian belief networks to health risk assessment, Stochastic Environmental Research and Risk Assessment 26 (3) (2012) 451–465.
[23] R. Lukman, D. Krajnc, P. Glavic, University ranking research, educational and environmental indicators, Journal of Cleaner Production 18 (2010) 619–628.
[24] D. Margaritis, S. Thrun, Bayesian Network Induction Via Local Neighborhoods. No. CMU-CS-99-134, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA, USA, 1999.
[25] C. Meek, Causal inference and causal explanation with background knowledge, in: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, 1995, pp. 403–410.
[26] J. Pearl, T. Verma, A theory of inferred causation, in: J. Allen, R. Fikes, E. Sandewall (Eds.), Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Morgan Kaufmann, New York, 1991, pp. 441–452.
[27] J.P. Pellet, A. Elisseeff, A partial correlation-based algorithm for causal structure discovery with continuous variables, in: 7th international Symposium on Intelligent Data Analysis VII, Springer, Berlin Heigelberg, 2007, pp. 229–239.
[28] J.P. Pellet, A. Elisseeff, Using Markov blankets for causal structure learning, The Journal of Machine Learning Research 9 (2008) 1295–1342.
[29] J. Rissanen, Fisher information and stochastic complexity, IEEE Transactions on Information Theory 42 (1996) 40–47.
[30] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, second ed., Prentice-Hall, Upper Saddle River, NJ, 2003.
[31] G. Schwarz, Estimating the dimension of a model, Annals of Statistics 6 (1978) 461–464.
[32] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, second ed., MIT Press, Cambridge, MA, 2000.
[33] I. Tsamardinos, C.F. Aliferis, A.R. Statnikov, Time and sample efficient discovery of Markov blankets and direct causal relations, in: The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 673–678.
[34] F.A. Vught, D. Westerheijden, Multidimensional ranking: a new transparency tool for higher education and research, Higher Education Management and Policy 23 (3) (2010) 32–56.
[35] A.V. Werhli, M. Grzegorczyk, D. Husmeier, Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks, Bioinformatics 22 (20) (2006) 2523–2531.
[36] Y. Zeng, X. He, Y. Xiang, H. Mao, Dynamic ordering-based search algorithm for Markov blanket discovery, in: Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, Berlin Heidelberg, 2011, pp. 420–431.