Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

European Journal of Operational Research 218 (2012) 230–238

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Computational Intelligence and Information Management

Exploring the critical pillars and causal relations within the NRI: An innovative approach

Wei-Wen Wu [a,*], Lawrence W. Lan [b,1], Yu-Ting Lee [a,2]

[a] Department of International Trade, Ta Hwa Institute of Technology, No. 1 Ta-Hwa Rd., Chiunglin, Hsinchu 30740, Taiwan
[b] Department of Television and Internet Marketing Management, Ta Hwa Institute of Technology, No. 1 Ta-Hwa Rd., Chiunglin, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 22 June 2011
Accepted 19 October 2011
Available online 31 October 2011

Keywords:
Data mining
Information and communication technologies (ICT)
Networked readiness index (NRI)
Causal relations
World Economic Forum

ABSTRACT

The *Global Information Technology Report* released by the World Economic Forum (WEF) has employed networked readiness index (NRI) to measure the global competitiveness of a country's information and communication technologies (ICT) diffusion. The final NRI overall scores were measured by an arithmetic mean aggregation of the composite pillars scores, which implicitly assumed that all the pillars have constant weights. The *Report* did not explore the critical pillars and causal relations for better decision making. To add values to this *Report*, the objective of this paper is to propose an innovative approach by using data mining techniques and partial least squares path modeling to scrutinize the critical pillars within the NRI and to further explore the causal relations amongst them. An empirical analysis based on the latest *Report (2009–2010)* is carried out. The results show that "business usage," "business readiness," and "market environment" are the three root drivers—critical pillars to manipulate the NRI overall scores; whereas "government readiness," which is further mostly affected by the "government usage," is the foremost enabler to the NRI overall scores. Based on the results, policy makers are suggested to allocate limited resources with priority to the three root drivers and one foremost enabler to frog-leap the global competitiveness of national ICT diffusion.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The World Economic Forum's *Global Information Technology Report Series* has raised awareness of the overall competitiveness of a nation's information and communication technologies (ICT) diffusion since its inaugural release in 2001. This *series* has identified the enabling factors that permit countries to fully benefit from ICT advances with emphasizing on the key responsibility from all actors, including individuals, businesses, and governments. The networked readiness index (NRI) was used as composite subindexes of environment, readiness, and usage in the latest *Report (2009–2010)*. To account for the three subindexes, nine pillars have been used, including E1 (market environment), E2 (political and regulatory environment), and E3 (infrastructure environment) under the environment subindex, R1 (individual readiness), R2 (business readiness), and R3 (government readiness) under the readiness subindex, U1 (individual usage), U2 (business usage), and U3 (government usage) under the usage subindex. A total of 68 components have been surveyed from each of the 133 countries to further account for the nine pillars. The final NRI overall scores were then computed as an arithmetic mean aggregation of the three composite subindex scores; whereas the score of each subindex was also an arithmetic mean aggregation of the scores of its corresponding composite pillars. In other words, the relation between the nine pillars and the NRI is assumed in ways that all the nine pillars have identical (with equal weights) and independent (no intertwining relations) contribution to the NRI overall scores. It suggests that if a country wishes to improve the NRI overall scores, ameliorating any of the nine pillars will contribute equally to the NRI overall scores. Besides, there was no information about the relations among the pillars, thus no ones will know about the most critical pillars (the root causes) to the NRI overall scores. While making decisions on allocating limited resources to accelerating a country's ICT global competitiveness, the policy makers may be curious about the interdependence among the pillars; more importantly, they may wish to know which pillar(s) are more critical than others in contributing to the NRI overall scores. Unfortunately, the *Report* did not explore any such causal information for better decision making.

Some recent studies addressed the same issue by arguing that the results of overall scores and rankings may greatly affected by the ways of selecting and weighting composite indicators for

* Corresponding author. Tel.: +886 3 5927700x2902; fax: +886 3 5925715.
E-mail addresses: itmike@thit.edu.tw (W.-W. Wu), lawrencelan@thit.edu.tw (L.W. Lan), ittina@thit.edu.tw (Y.-T. Lee).
[1] Addresses: National Chiao Tung University, Taiwan; Chinese Institute of Business Innovation and Development, Taiwan. Tel.: +886 3 5927700x3210; fax: +886 2 23375755.
[2] Tel.: +886 3 5927700x2902; fax: +886 3 5925715.

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.10.013

Author's personal copy

national policies. For instance, Grupp and Mogee (2004) found that the composite scores and country rank positions of science and technology (S&T) indicators can vary considerably depending on the selection process; thus, the use of scoreboards leaves room for manipulation in the policymaking system. Cerulli and Filippetti (2010) employed a sensitivity analysis to detect the subindexes importance. Grupp and Schubert (2010) suggested subindex weights be chosen on the basis of shadow prices, rather than by equal weighting or other automatic methods. It would be more reasonable if one could have objectively reflected the relative importance of a set of criteria (pillars). Moreover, the final NRI overall scores and rankings have revealed no information about the causal relations between the pillars and between the pillars and the NRI overall scores. It would be more useful if one could have further formulated causal relations amongst the pillars for national policies. With causal information, the policymakers can concentrate on the root drivers—the most critical causal pillars and their associated components to transform strategic objectives into effective actions. However, little has been found to explore the causal relations between the pillars and between the pillars and the NRI overall scores for better national policy making.

To fill up these gaps, this paper aims to propose an innovative approach that can establish the causal relations amongst the pillars without subjectively assigning an equal weight to them, make use of the causal directions to create hypotheses, and perform the hypotheses test to explore the most critical pillars that have mostly affected the national overall ICT competiveness. The remainder of this paper is organized as follows. Section 2 explains the proposed approach together with the three techniques used—expectation maximization (EM) clustering, Bayesian network (BN) classifiers with tree augmented NaïveBayes (TAN) search algorithm, and partial least squares (PLS) path modeling. Section 3 conducts the empirical analysis by implementing the proposed approach. Section 4 discusses the policy implications based on the results. Finally, the conclusions, limitations and recommendations for future research are presented.

2. Methodology

This paper presents a new perspective to elucidate the well-known ICT index—NRI. Our goal is not to propose a new way to compute this index but to better explain the relations between subindexes/pillars and between subindexes/pillars and the obtained NRI overall scores. Identifying the critical pillars within the NRI and exploring the causal relations amongst them is a complicated and challenging issue. Basically, to perform causal analyses, one needs to investigate causal directions amongst the subindexes/pillars. Once the causal directions are confirmed, one can then create the hypotheses accordingly. By performing the hypotheses test, one can further scrutinize the most critical subindexes/pillars that have significantly affected the country's overall ICT competiveness. Based on this rationale, an innovative approach is proposed as follows.

2.1. The proposed approach

The proposed approach contains four steps as depicted in Fig. 1.

Step 1: Cluster the ranked countries based on the NRI rankings and scores. This study adopts the EM clustering technique.
Step 2: Create a causal diagram which displays the causal directions amongst the pillars and NRI overall scores. This study employs the BN classifier with the TAN search algorithm.
Step 3: Create the hypotheses by referring to the resulted causal diagram.

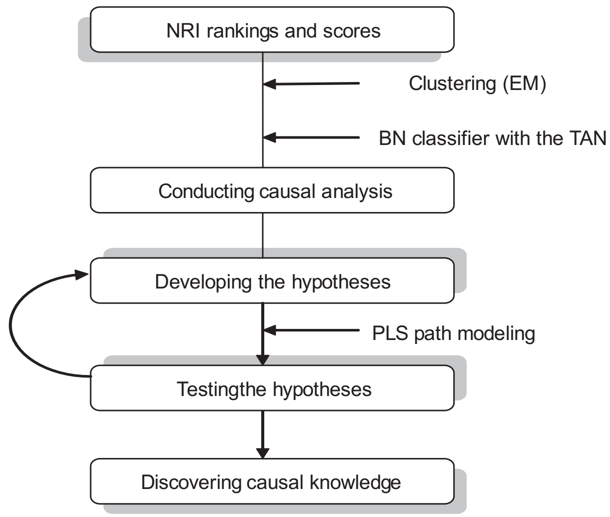

Fig. 1. The proposed approach.

Step 4: Test the hypotheses to confirm the significant paths. This study employs the PLS path modeling.

Three important techniques are employed in different steps of the proposed approach: EM clustering, BN-TAN data mining (BN classifier with the TAN search algorithm), and PLS-PM (path modeling). The reasons for choosing these three techniques are briefly explained as follows: (1) EM clustering can automatically group the ranked nations into appropriate classes (a categorical variable) without subjectively assuming the number of classes before the clustering or assigning weights to the pillars. The categorical variable is a prerequisite for conducting the BN-TAN data mining in the next step. (2) BN-TAN data mining has been viewed as a powerful technique to create the causal diagram, based on which we can further develop the hypothesized causal relations between the pillars and between the pillars and the NRI overall scores. (3) PLS-PM is a powerful and widely used technique for validating the hypotheses. We employ this technique to confirm the significant paths by testing the hypotheses developed in the previous step. The rationales or advantages of these three techniques are elaborated in the subsequent sections.

2.2. EM clustering

The goal of clustering is to assign the objects/instances into clusters which are grouped with similar characteristics (Nascimento and de Carvalho, 2011; Fernandez et al., 2010). Clustering analysis has already been used in many application domains. It is particularly important if one wishes to conduct the task of data preprocessing for statistical analyses (e.g. discriminant analysis) and data mining applications (e.g. classification, association analysis). Karaboga and Ozturk (2011) indicated that (1) clustering is the process of using similarity measures to distinguish natural clusters in multidimensional data; (2) clustering techniques can be roughly divided into unsupervised clustering and supervised clustering; and (3) unlike supervised clustering, unsupervised clustering does not need to specify the number of clusters/classes beforehand. Unsupervised clustering is also named as automatic clustering. Although many clustering techniques have been developed, most of them require that the users know the number of clusters/classes. In fact, it is difficult to know the number of clusters when the dataset have high dimension.

The EM clustering algorithm assigns a probability distribution to each instance which indicates the probability of it belonging

Author's personal copy

to each of the clusters. Importantly, the EM clustering algorithm can select the number of clusters automatically by cross-validation. According to Witten and Frank (2005), for the implementation of EM clustering, users can specify how many clusters to generate or use the mode of 10-fold cross-validation. WEKA (Waikato Environment for Knowledge Analysis) is a useful suite of machine learning software. It provides the EM clustering algorithm that can be performed as the automatic clustering. Details of the EM algorithm can be referred to Witten and Frank (2005).

2.3. BN classifier

The BN is a graphical representation of probabilistic relationships between multiple attributes/variables (Nadkarni and Shenoy, 2001). It is a directed acyclic graph (DAG) that consists of a set of nodes/vertices linked by arcs, in which the nodes represent the attributes and the arcs stand for relationships among the connected attributes (Hruschka and Ebecken, 2007). BN incorporates probabilistic inference engines that support reasoning under uncertainty (Hruschka and Ebecken, 2007). It is an outcome of a machine-learning process that finds a given network's structure and its associated parameters, and it can provide diagnostic reasoning, predictive reasoning, and intercausal reasoning (Lauria and Duchessi, 2007).

Inferring Bayesian structure from expression data can be viewed as a search problem in the network space (Wang et al., 2004). Thus, to search the BN space, it is necessary to employ a variety of search methods, such as Simulated Annealing (SA) algorithm, Genetic Algorithm (GA), and Tree Augmented NaïveBayes (TAN). TAN is an extension of NaïveBayes—it removes the NaïveBayes assumption that all the attributes are independent (Baesens et al., 2004). More importantly, TAN can produce a causal-effect graph (not just a tree-like graph), in which the only and greatest parent node for all other nodes is located at the top in the DAG (Friedman et al., 1997). The causal-effect graph of TAN is formed by calculating the maximum weight spanning tree using Chow–Liu method (Chow and Liu, 1968). The BN classifier incorporated in WEKA, such as BN with TAN search algorithm, has exhibited excellent performance in data mining (Cerquides and De Mantaras, 2005). Details of BN classifier with TAN algorithm (BN-TAN) can be referred to Friedman et al. (1997).

2.4. PLS path modeling

It is well known that both covariance-based structural equation modeling (SEM) and partial least squares (PLS) path modeling are the two main approaches to establishing the relationships between latent variables (e.g., Curkovic, 2003; Tenenhaus et al., 2005; Temme et al., 2006; Sohn et al., 2007; Su and Yang, 2010; Wu, 2010). Covariance-based SEM requires assumption of homogeneity in the observed population. It focuses on maximizing the explained covariation among various constructs with theory confirmation. Thus, covariance-based SEM is likely a preferred method if the objective is only a description of theoretical constructs with no interest in inference to observable variables (Anderson and Vastag, 2004). In contrast, PLS path modeling requires minimal assumptions about the statistical distributions of data sets. It maximizes the explained variation among various constructs with emphasis on the causal explanation, which is particularly advantageous when models are complex (Fornell and Bookstein, 1982; Hulland, 1999; Lauria and Duchessi, 2007). More importantly, PLS path modeling can work well with small sample size, say, 30 (Wixom and Watson, 2001; Ranganathan and Sethi, 2002; Anderson and Vastag, 2004; Lee et al., 2009). Therefore, the proposed approach employs the PLS path modeling to test the hypotheses. Details of the PLS path modeling can be referred to Jakobowicz and Derquenne (2007) or Henseler et al. (2009).

3. Empirical analysis

Based on the NRI overall scores and rankings presented in the WEF's *Annual Report (2009–2010)* an empirical analysis is further conducted by implementing the four steps of the proposed approach. The following presents the detailed results.

3.1. The clustering

In step 1, it requires identifying the explanatory and outcome variables to group the ranked countries into proper classes. The EM clustering algorithm is employed in which the nine pillars are used as the explanatory variables and the resulted class is used as the outcome variable. Table 1 presents the results of EM clustering. It shows that the 133 ranked countries can be automatically grouped into four classes without specifying the number of classes in advance or assigning subjective weights to the pillars subjectively. As a result, the 133 ranked countries have been automatically grouped into four classes—a categorical variable readily for use in BN-TAN data mining in the next step. The geometric mean of the Overall Score in each class reveals that class 3 (the highly developed ICT group) has mean Overall Score of 5.28, followed by class 1 (4.38), class 2 (3.81), and the least developed ICT group—class 4 (3.07).

We further validate the classification accuracy by different popular classification algorithms and the results are presented in Table 2. It shows that the EM clustering results are quite satisfactory because the mean accuracy has reached as high as 92.53%, based on the geometric mean of different classification algorithms.

3.2. The causal diagram

In step 2, we employ BN classifier with TAN search algorithm, using the overall scores of nine pillars as the inputs and the categorical results of Cluster-EM as the output to conduct the causal analysis. It is implemented with the software WEKA using a test mode of 10-fold cross-validation. Fig. 2 displays the causal diagram, from which one can obviously see the causal directions between pillars and Cluster-EM.

3.3. The initial hypotheses

In step 3, we create the initial hypotheses based on Fig. 2. However, the causal directions acquired by BN classifier with the TAN search algorithm must reverse its directions before one can apply the PLS path modeling (Wu, 2010, 2011). Verma and Pearl (1991) also remarked that if any two causal models are statistically equivalent, the causal directions can be reversed. Accordingly, all the hypothesized causal directions are reversed as shown in Fig. 3. It indicates that a total of seventeen hypotheses can be created—nine hypotheses showing the direct effect of each pillar on the Overall Score and eight hypotheses relating to each other between the pillars. To name one example, one hypothesis is that U2 (business usage) will positively affect not only U3 (government usage) but also Overall Score.

3.4. Testing the hypotheses

In step 4, we test the above 17 initial hypotheses by using the PLS path modeling method, which is implemented with the software SmartPLS. After removing the insignificant paths, the significant paths among pillars and Overall_Score are displayed in Fig. 4.

Author's personal copy

Table 1
The results of EM clustering.

Country/Economy	El	E2	E3	Rl	R2	R3	Ul	U2	U3	Overall Score	Class
Bahrain	4.71	4.37	3.28	5.70	4.32	4.75	2.90	4.77	4.58	4.38	1
Barbados	3.90	5.09	4.02	5.84	4.50	4.48	3.38	4.54	3.65	4.38	1
Chile	4.43	4.74	3.07	5.45	4.71	4.42	2.49	5.06	4.53	4.32	1
China	4.01	4.50	2.71	5.66	4.65	4.58	1.73	4.69	4.77	4.15	1
Cyprus	4.31	4.99	3.95	6.07	4.57	4.56	3.28	4.86	4.07	4.52	1
Czech Republic	4.32	4.30	3.78	6.02	5.15	4.42	3.91	5.32	3.53	4.53	1
Hungary	4.17	4.33	3.78	5.58	4.64	4.03	3.44	4.69	3.86	4.28	1
India	4.12	4.19	2.70	5.57	5.05	4.12	1.26	5.09	4.19	4.03	1
Jordan	4.12	4.70	3.01	5.52	4.19	4.72	1.93	5.02	4.46	4.19	1
Lithuania	4.14	4.57	3.58	5.67	4.56	4.29	3.51	4.74	4.55	4.40	1
Mala	4.47	4.99	3.91	5.95	4.64	5.32	3.47	4.95	5.37	4.79	1
Mauritius	4.65	4.62	2.68	5.53	4.40	4.34	2.15	4.46	3.79	4.07	1
Oman	4.31	4.52	2.69	5.54	4.60	4.47	1.97	4.44	4.21	4.08	1
Portugal	4.41	4.85	3.76	5.55	4.72	5.16	3.07	5.17	4.98	4.63	1
Puerto Rico	4.49	4.75	3.51	5.42	5.12	3.77	2.05	5.02	3.97	4.23	1
Qatar	4.72	4.74	3.76	6.05	5.01	4.99	3.26	4.86	4.73	4.68	1
Saudi Arabia	4.35	4.55	3.44	5.06	5.04	4.55	2.38	4.87	4.30	4.28	1
Sbvak Republic	4.33	4.17	3.23	5.65	4.74	3.98	3.33	4.91	3.31	4.19	1
Slovenia	4.19	4.32	4.18	5.94	5.02	4.45	3.68	4.98	4.37	4.57	1
South Africa	4.44	5.04	2.83	5.02	4.88	3.99	1.83	4.76	3.79	4.07	1
Spain	4.16	4.60	3.83	5.66	5.12	4.50	3.57	4.85	4.23	4.50	1
Thailand	4.42	4.43	2.97	5.52	4.67	4.22	1.98	4.70	4.32	4.14	1
Tunisia	4.16	4.88	3.20	5.94	4.91	4.78	1.87	4.97	4.36	4.34	1
United States	4.74	4.62	3.52	5.89	4.94	5.33	3.57	5.18	5.06	5.68	1
Geometric mean	4.33	4.61	3.36	5.65	4.75	4.49	2.63	4.87	4.26	4.38	
Argentina	3.04	3.21	3.12	5.25	4.48	3.25	2.38	4.18	3.35	3.58	2
Azerbaijan	3.97	4.22	2.66	5.06	4.36	4.46	1.57	4.61	4.42	3.93	2
Botswana	3.91	4.44	2.70	5.31	3.94	3.97	1.65	3.98	3.60	3.72	2
Brazil	3.28	3.81	2.91	5.02	4.83	3.98	2.12	5.06	4.47	3.94	2
Brunei Darussalam	3.45	3.87	2.65	5.64	4.27	4.38	2.38	4.41	3.76	3.87	2
Bulgaria	3.77	3.69	3.18	5.34	4.11	3.93	2.74	3.97	3.50	3.80	2
Colombia	3.59	3.87	2.77	5.37	4.50	4.28	2.03	4.36	4.06	3.87	2
Costa Rica	4.02	3.85	3.13	5.57	4.91	4.30	2.10	4.26	3.77	3.99	2
Croatia	3.82	4.04	3.39	5.62	4.64	4.13	3.01	4.57	3.55	4.09	2
Dominican Republic	3.97	3.90	2.22	4.87	4.00	4.31	1.70	4.54	4.38	3.76	2
Egypt	4.00	4.12	2.78	4.66	3.86	4.27	1.49	4.64	4.00	3.76	2
El Salvador	4.04	3.75	2.15	4.99	4.14	3.90	1.87	4.34	4.03	3.69	2
Greece	3.79	4.13	4.10	5.49	4.27	3.97	2.51	4.28	3.41	4.00	2
Guatemala	4.05	3.52	2.05	4.81	4.26	3.79	1.69	4.80	3.78	3.64	2
Indonesia	4.05	3.80	2.31	5.52	4.58	3.43	1.34	4.44	3.15	3.62	2
Italy	3.85	3.71	3.70	5.47	4.78	3.77	3.84	4.61	3.69	4.16	2
Jamaica	3.97	4.21	2.71	5.14	4.29	3.97	3.67	4.49	3.86	4.03	2
Kazakhstan	3.73	4.02	3.15	4.73	4.51	4.14	2.01	4.05	3.73	3.79	2
Kuwait	4.44	4.19	3.44	5.37	4.30	3.80	2.41	4.63	3.23	3.98	2
Latvia	4.12	4.45	3.36	5.52	4.38	3.88	3.18	4.49	3.57	4.10	2
Macedonia, FYR	3.75	3.64	2.92	5.14	4.16	3.97	2.42	3.69	3.34	3.67	2
Mexico	3.88	3.77	2.78	5.12	4.26	3.99	2.04	4.31	4.39	3.84	2
Montenegro	3.93	3.72	3.84	4.81	3.98	3.83	2.96	3.94	3.11	3.79	2
Morocco	4.03	3.93	2.62	4.89	3.96	3.54	1.85	4.21	3.30	3.59	2
Panama	4.17	4.05	2.87	5.19	4.21	4.01	1.80	4.57	3.68	3.84	2
Peru	3.69	3.45	2.29	4.68	4.11	3.59	1.95	4.09	3.35	3.47	2
Philippines	4.09	3.74	2.26	4.98	3.91	3.75	1.61	4.53	3.51	3.60	2
Poland	3.71	3.59	3.51	5.57	4.56	3.54	2.79	4.36	2.57	3.80	2
Romania	3.90	4.04	2.95	5.55	4.47	3.99	3.16	4.28	3.42	3.97	2
Russian Federation	3.58	3.76	3.63	5.39	4.26	3.79	2.33	4.23	2.95	3.77	2
Senegal	3.99	3.59	2.40	4.59	4.28	3.95	1.35	4.67	4.21	3.67	2
Serbia	3.38	3.64	3.17	5.44	4.18	3.77	2.23	3.57	3.19	3.62	2
Sri Lanka	4.00	4.00	2.66	5.09	4.10	4.22	1.39	4.84	3.78	3.79	2
Trinidad and Tobago	3.91	3.36	2.89	5.55	4.24	3.65	2.24	4.10	3.04	3.67	2
Turkey	3.97	4.20	2.94	5.35	4.31	3.75	2.13	4.74	3.79	3.91	2
Ukraine	3.60	3.67	3.50	5.52	4.16	3.90	2.22	4.35	3.96	3.88	2
Uruguay	3.52	4.18	2.88	5.35	4.18	4.16	2.36	4.32	3.68	3.85	2
Vietnam	3.72	4.23	2.66	4.88	4.25	4.42	1.62	4.23	4.12	3.79	2
Geometric mean	3.82	3.87	2.89	5.20	4.28	3.93	2.14	4.35	3.62	3.81	
Australia	4.79	5.81	5.06	6.18	5.24	5.14	4.47	5.40	5.55	5.29	3
Austria	4.77	5.93	4.28	6.24	5.55	4.94	4.26	5.87	5.18	5.22	3
Belgium	4.69	5.24	4.44	6.32	5.67	4.55	4.37	5.55	4.34	5.02	3
Canada	5.07	5.52	5.57	6.26	5.31	5.03	4.93	5.77	5.23	5.41	3
Denmark	5.14	6.03	5.35	6.32	5.72	5.85	5.96	6.15	6.09	5.85	3
Estonia	4.85	5.26	4.01	6.06	4.94	5.44	4.83	5.29	6.00	5.19	3
Finland	5.35	5.97	5.20	6.54	5.78	5.38	4.59	6.04	4.92	5.53	3
France	4.72	5.53	4.48	6.11	5.51	5.03	4.13	5.65	5.40	5.17	3
Germany	4.88	5.83	4.56	5.97	5.78	4.70	4.51	6.04	4.25	5.17	3

(*continued on next page*)

Author's personal copy

Table 1 (*continued*)

Country/Economy	El	E2	E3	Rl	R2	R3	Ul	U2	U3	Overall Score	Class
Hong Kong SAR	5.61	5.79	3.68	6.27	5.09	5.02	5.17	5.46	5.64	5.30	3
Iceland	5.10	5.80	6.02	6.30	5.27	5.28	4.94	5.91	4.86	5.50	3
Ireland	5.16	5.53	4.56	6.04	5.63	4.66	3.95	5.14	4.59	5.03	3
Israel	5.00	4.74	4.51	5.80	5.42	5.11	3.67	5.78	4.75	4.98	3
Japan	5.10	5.52	4.29	5.90	5.60	4.90	4.85	6.09	4.43	5.19	3
Korea, Rep.	5.08	5.20	4.79	6.22	5.40	5.70	4.31	5.70	5.95	5.37	3
Luxembourg	5.02	5.59	3.84	5.95	4.78	5.05	5.69	5.29	4.64	5.10	3
Malaysia	4.92	5.17	3.29	5.97	5.25	5.20	2.86	5.26	4.89	4.76	3
Netherlands	5.15	5.78	4.68	6.11	5.63	4.98	5.39	5.74	4.82	5.48	3
New Zealand	4.59	5.58	5.02	6.04	4.98	4.59	4.25	5.23	5.09	5.04	3
Norway	4.90	5.87	5.40	6.09	5.33	5.49	5.18	5.85	5.33	5.49	3
Singapore	5.58	6.30	4.13	6.50	5.62	5.92	5.21	5.77	6.02	5.67	3
Sweden	5.11	5.92	5.73	6.26	5.86	5.72	5.06	6.15	5.72	5.84	3
Switzerland	5.43	5.85	5.05	6.43	6.00	5.07	5.29	6.11	5.02	5.58	3
Taiwan, China	5.43	4.54	4.96	6.22	5.49	5.14	4.60	5.74	5.55	5.30	3
United Arab Emirates	5.59	5.45	5.72	6.14	5.83	5.46	5.05	6.06	5.79	4.76	3
United Kingdom	5.09	5.45	4.82	6.02	5.38	4.88	5.45	5.65	4.73	5.27	3
Geometric mean	5.07	5.57	4.70	6.16	5.46	5.15	4.74	5.71	5.16	5.28	
Albania	3.42	3.42	2.09	4.79	3.34	3.77	.74	3.46	3.06	3.23	4
Algeria	3.04	3.20	2.56	4.51	3.84	3.44	.74	3.22	2.75	3.14	4
Armenia	3.27	3.39	2.62	3.88	3.52	3.27	.33	3.63	2.66	3.06	4
Bangladesh	3.46	2.86	1.96	2.88	3.09	3.06	.17	3.32	2.53	2.70	4
Benin	3.38	3.72	2.12	2.58	3.16	3.53	.16	3.49	3.51	2.96	4
Bolivia	2.80	2.87	2.54	3.75	3.34	2.79	.41	3.09	2.83	2.82	4
Bosnia and Herzegovina	3.18	3.20	2.57	5.13	3.58	3.22	.95	3.70	2.56	3.23	4
Burkina Faso	3.59	3.95	2.12	2.34	3.55	3.49	.08	3.79	3.70	3.07	4
Burundi	2.77	2.94	1.95	2.80	3.03	2.92	.03	3.33	2.92	2.63	4
Cambodia	3.40	3.38	1.71	2.93	3.22	3.47	.12	3.61	3.15	2.89	4
Cameroon	3.09	3.19	1.85	3.48	3.54	3.23	.20	3.70	3.07	2.93	4
Chad	2.54	2.75	1.63	2.47	2.85	2.75	.06	3.11	2.79	2.44	4
Cote d'Ivoire	3.46	3.11	2.52	3.61	3.74	3.23	.29	4.31	2.76	3.12	4
Ecuador	3.12	2.96	1.94	4.81	3.72	2.97	.92	3.29	2.56	3.03	4
Ethiopia	3.33	3.51	1.99	2.57	3.01	3.30	.01	3.50	2.96	2.80	4
Gambia, The	3.59	4.65	1.90	3.05	3.95	4.11	.42	4.32	4.01	3.44	4
Georgia	4.07	3.71	2.49	5.08	3.45	3.83	.46	3.85	3.36	3.48	4
Ghana	3.75	4.06	2.25	3.82	3.99	3.62	.26	3.56	2.97	3.25	4
Guyana	3.44	3.42	2.62	4.77	3.86	3.56	.65	3.46	2.81	3.29	4
Honduras	3.80	3.67	2 22	4.66	3.95	3.56	.28	4.10	3.41	3.41	4
Kenya	3.60	3.74	2.63	3.36	4.16	3.75	.33	4.23	3.37	3.35	4
Kyrgyz Republic	3.37	3.50	2.60	4.17	3.16	3.07	.49	3.43	2.61	3.04	4
Lesotho	3.55	3.44	2.60	3.55	3.17	3.34	.19	3.49	2.86	3.02	4
Libya	3.02	3.42	3.03	4.77	3.78	3.46	.59	3.51	2.95	3.28	4
Madagascar	3.55	3.43	1.98	3.31	3.45	3.82	.08	3.79	3.36	3.09	4
Malawi	3.59	3.93	2.21	3.29	3.94	3.65	.06	3.46	2.91	3.12	4
Malr	3.51	3.88	2.14	2.70	3.20	3.91	.15	4.12	3.99	3.18	4
Mauritania	3.57	3.44	1.71	3.27	3.67	3.56	.35	3.96	3.58	3.12	4
Moldova	3.29	3.89	2.67	3.98	3.22	3.59	.80	4.14	3.10	3.30	4
Mongolia	3.63	3.56	2.74	4.68	3.46	4.07	.52	3.71	3.50	3.43	4
Mozambique	3.35	3.59	1.91	2.53	3.20	3.39	.12	3.67	3.47	2.91	4
Namibia	3.89	4.74	2.39	4.35	3.88	3.22	.57	4.19	2.70	3.44	4
Nepal	3.77	3.41	1.91	3.43	3.23	3.01	.04	3.37	2.49	2.85	4
Nicaragua	3.42	3.24	2.01	3.89	2.93	3.10	.35	3.44	2.71	2.90	4
Nigeria	3.91	4.06	2.01	4.21	4.53	3.22	.28	4.39	3.45	3.45	4
Pakistan	3.75	3.46	1.99	3.79	4.14	3.64	.59	4.16	3.27	3.31	4
Paraguay	3.57	2.77	2.43	4.23	3.34	2.78	.67	3.12	2.49	2.93	4
Suriname	3.16	2.74	2.43	4.92	4.26	2.67	.71	3.31	2.06	3.03	4
Syria	3.53	3.53	2.48	5.11	4.06	3.70	.54	3.89	2.83	3.41	4
Tajikistan	3.10	4.02	2.57	4.21	3.60	3.73	.04	3.84	3.11	3.25	4
Tanzania	3.45	3.81	1.96	3.14	3.46	3.47	.15	3.72	2.93	3.01	4
Timor-Leste	3.24	2.47	1.74	2.36	2.64	2.96	.05	2.94	2.81	2.47	4
Uganda	3.41	3.66	2.27	1.95	3.19	3.65	.19	3.97	3.54	2.98	4
Venezuela	2.67	2.94	2.92	5.13	4.01	3.44	2.08	3.92	3.39	3.39	4
Zambia	3.84	4.07	2.21	3.97	3.79	3.29	.21	3.87	3.10	3.26	4
Zimbabwe	2.69	3.02	2.36	2.49	3.31	2.44	.27	2.87	2.00	2.49	4
Geometric mean	3.37	3.44	2.23	3.60	3.51	3.35	.34	3.64	2.99	3.07	

The significant path coefficients are detailed in Table 3. From Fig. 4, we note that (1) the highest path coefficient (0.886) is the E2 (political and regulatory environment) → R3 (government readiness); (2) the U3 (government usage) exhibits the best ability to explain this model (R^2 = 88.6%); and (3) the combination of these nine pillars has predictive ability of 98% for the Overall_Score. We note that three path coefficients (R1 → Overall_Score, R2 → Overall_Score, and U2 → U3) are less than 0.20, suggesting that these relations may not permit meaningful interpretations, according to Chin (1998).

It should be noted that one can further refine the hypotheses on the basis of the initial hypotheses developed by Fig. 3. We attempt to improve the initial hypotheses by assuming direct effects from E1 (market environment) to other pillars as shown in Fig. 5. The

Author's personal copy

Table 2
The classification accuracy of EM clustering.

Classification algorithm	Classification accuracy (%)
NaiveBayes	99.25
BN-TAN	91.79
Logistic (multinomial logistic regression)	90.30
MLP (Multilayer Perception)	97.01
RBFNetwork	98.51
IBk (K-nearest neighbors)	94.78
J48 (the Weka version of C4.5)	83.58
Geometric mean	92.53

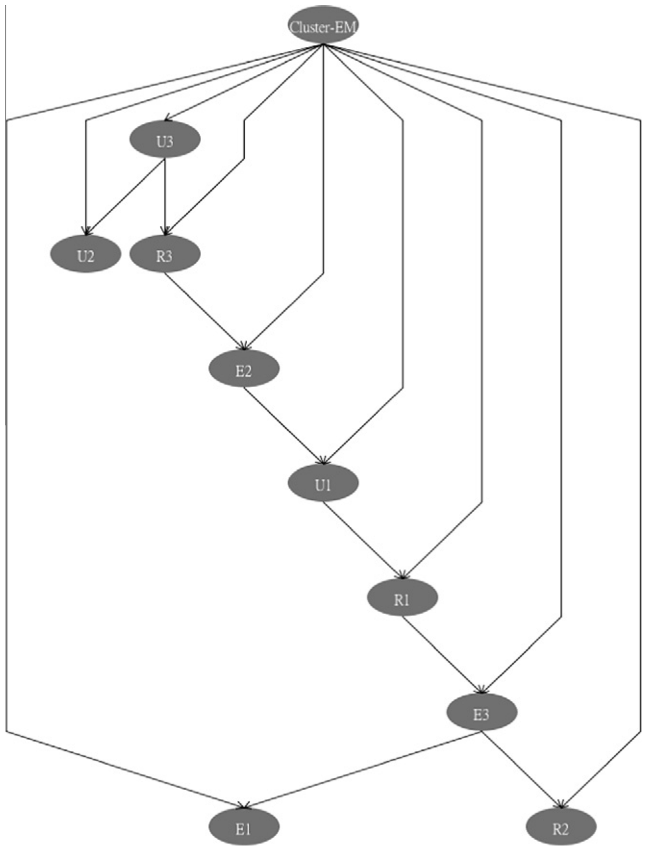

Fig. 2. Causal diagram acquired by BN classifier.

corresponding refined results are presented in Table 4. From Fig. 5, we note that (1) the highest path coefficient (0.666) is the (market environment) → E2 (political and regulatory environment); (2) the R3 (government readiness) displays the best ability to explain this model (R^2 = 91.2%); and (3) the combination of these nine pillars has predictive ability of 98% for the Overall Score. We also note that two path coefficients (R1 → Overall_Score, R2 → Overall_Score) are still less than 0.20, but the best ability to explain the model (R^2 = 91.2%) has been improved and the number of low path coefficients (less than 0.20) has been reduced from three to two. In other words, the refined hypotheses can lead to slightly better results than the initial hypotheses; therefore, the following discussion will be based on the refined results.

4. Discussion and implications

We discover the causal knowledge and draw some policy implications based on the refined results. First of all, according to the *2002–03 edition of Global Information Technology Report*, the NRI was based on three main principles: (1) environment was the crucial enabler of networked readiness, (2) a multi-stakeholder effort was required, and (3) ICT readiness facilitated ICT usage. The fundamental logic followed a three-stage value chain "ICT environment → ICT readiness → ICT usage" featured by "antecedent and subsequent" with an input–output relationship as well as with a simple linear correlation. One may presume that the logic of three-stage value chain is perhaps reasonable, but the ICT diffusion has been more complex than one can envisage. The empirical results of this study have evidenced that the relationships among the three stages do have the input–output connection, but they are not with a simple linear correlation. Both Figs. 4 and 5 have evidenced that the ICT diffusion is actually affected by the way of nine pillars' interactions, rather than a simple three-stage value chain.

Further causal knowledge can be discovered from Fig. 5. Firstly, U2 (business usage) positively affects not only Overall Score but also U3 (government usage), which in turn influences R3 (government readiness), and that has a positive effect on Overall Score. Secondly, R2 (business readiness) positively affects not only Overall Score but also E3 (infrastructure environment), which in turn influences the path "R1 (individual readiness) → U1 (individual usage) → E2 (political and regulatory environment) → R3 (government readiness) → Overall_Score." Thirdly, E1 (market environment) has positive direct and indirect effects on many pillars (U3, R3, E3, R1, U1, E2) and Overall Score.

From the above, one can learn that U2 (business usage), R2 (business readiness) and E1 (market environment) can be viewed as the "root drivers" to manipulate the Overall Score. E3 (infrastructure environment) and R3 (government readiness) are sensitively affected by many other pillars. In particular, R3 (government readiness) (R^2 = 91.2%) is the foremost enabler as the central component of NRI, yet it is influenced by a series of antecedents.

Based on the findings, some policy implications can be derived. For example, it is sensible to focus on three specific pillars (U2, R2 and E1) rather than all nine pillars. Accordingly, U2, R2 and E1 should be treated as the mostly imperative catalysts if the policymakers wish to promote firms' entrepreneurship and innovation in ICT industry, as well as to advance a sound and high-quality environment for ICT market. As for E3 (infrastructure environment), the improved ICT infrastructure is dependent on R2 (business readiness) and E1 (market environment), of which R2 with path coefficient (0.636) is more crucial than E1 (0.244). As for the improvement of R3 (government readiness), apart from making effort to boost ICT readiness of government, policymakers need to facilitate several antecedent pillars—in particular, U3 (government usage) with high path coefficient (0.621) is the foremost pillar and it has positive direct effect on R3.

In sum, the policy makers are suggested to allocate limited resources with top priority to the three root drivers—"business usage," "business readiness," and "market environment" and one foremost enabler—"government usage" to frog-leap the global competitiveness of national ICT diffusion.

5. Conclusions

In order to increase the credibility and to make use of the NRI scores and rankings from the World Economic Forum's *Annual Report*, this paper has proposed an innovative approach to properly formulating the causal relations among nine pillars and NRI overall score, to creating and testing the hypotheses so that the most critical ones can be scrutinized. The proposed approach has employed several techniques: EM clustering, BN classifier with TAN search algorithm, and PLS path modeling. Several conclusions have been reached as follows. In lieu of focusing on all nine pillars simultaneously, the policymakers may spotlight on U2 (business usage), R2 (business readiness) and E1 (market environment) with priority

Author's personal copy

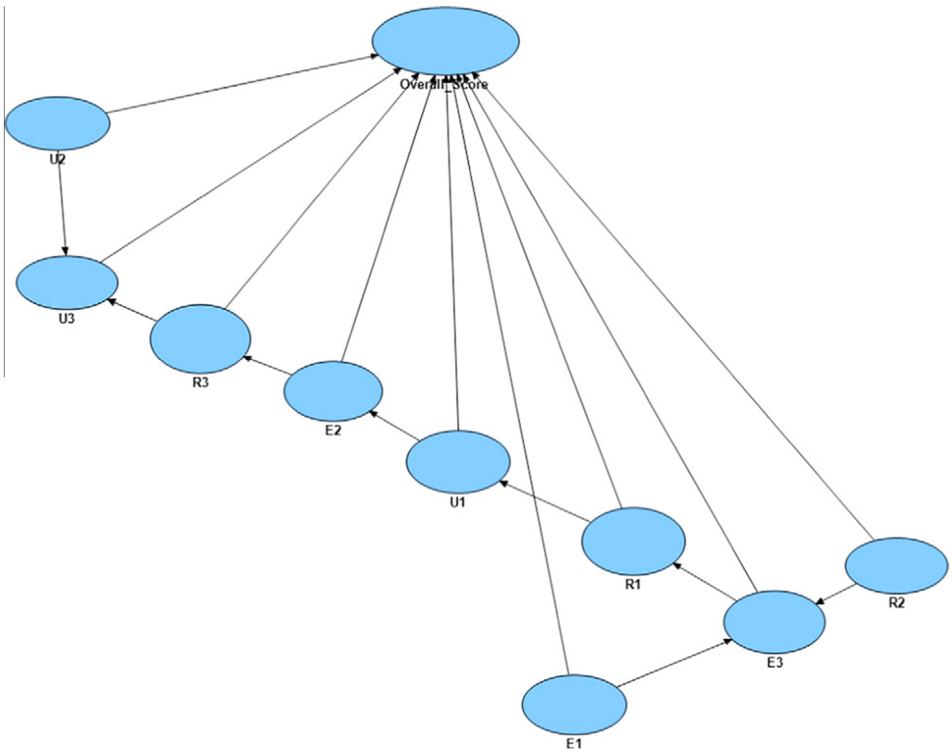

Fig. 3. Initial causal relations among pillars and Overall_Score.

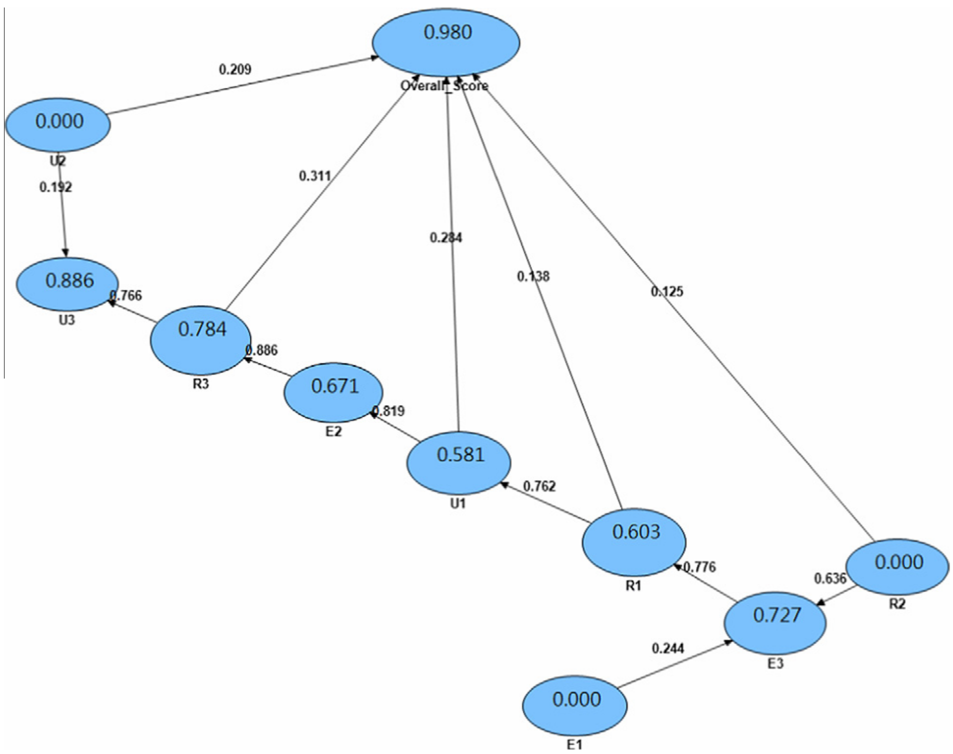

Fig. 4. The initial significant paths among pillars and Overall_Score.

because these three pillars have served as the "root drivers" to manipulate the NRI overall scores. R3 (government readiness) is the foremost enabler—it is affected by many antecedents, especially by U3 (government usage). Therefore, it is favorable to improve U2, R2 and E1, in the meantime to advance U3 (government usage) in order to effectively boost the national ICT competitiveness.

Nowadays, ICT has proven to empower individuals with unprecedented access to information and knowledge, with important consequences in providing education, doing business, linking social interactions, among others. ICT has also played a formidable role in reducing poverty, improving living conditions, and providing opportunities for the poor in ways to help sustain a country's

Author's personal copy

Table 3
The coefficients of initial significant paths.

	Original Sample (O)	Sample Mean (M)	Standard deviation (STDEV)	Standard error (STERR)	*T* statistics (|O/STERR|)
El → E3	0.24433	0.24502	0.07826	0.07826	3.12193
E2 → R3	0.88557	0.88530	0.02022	0.02022	43.79709
E3 → Rl	0.77630	0.77985	0.02521	0.02521	30.79351
Rl → Overall_Score	0.13803	0.13800	0.02618	0.02618	5.27281
Rl → Ul	0.76205	0.76642	0.02570	0.02570	29.64952
R2 → E3	0.63592	0.63692	0.07472	0.07472	8.51127
R2 → Overall_Score	0.12456	0.12298	0.03661	0.03661	3.40229
R3 → Overall_Score	0.31079	0.31536	0.04357	0.04357	7.13269
R3 → U3	0.76601	0.77260	0.07698	0.07698	9.95069
Ul → E2	0.81937	0.82095	0.03278	0.03278	24.99458
Ul → Overall_Score	0.28387	0.28499	0.02179	0.02179	13.03030
U2 → Overall_Score	0.20907	0.20444	0.03845	0.03845	5.43794
U2 → U3	0.19225 1	0.18511	0.08279 1	0.08279 1	2.32218

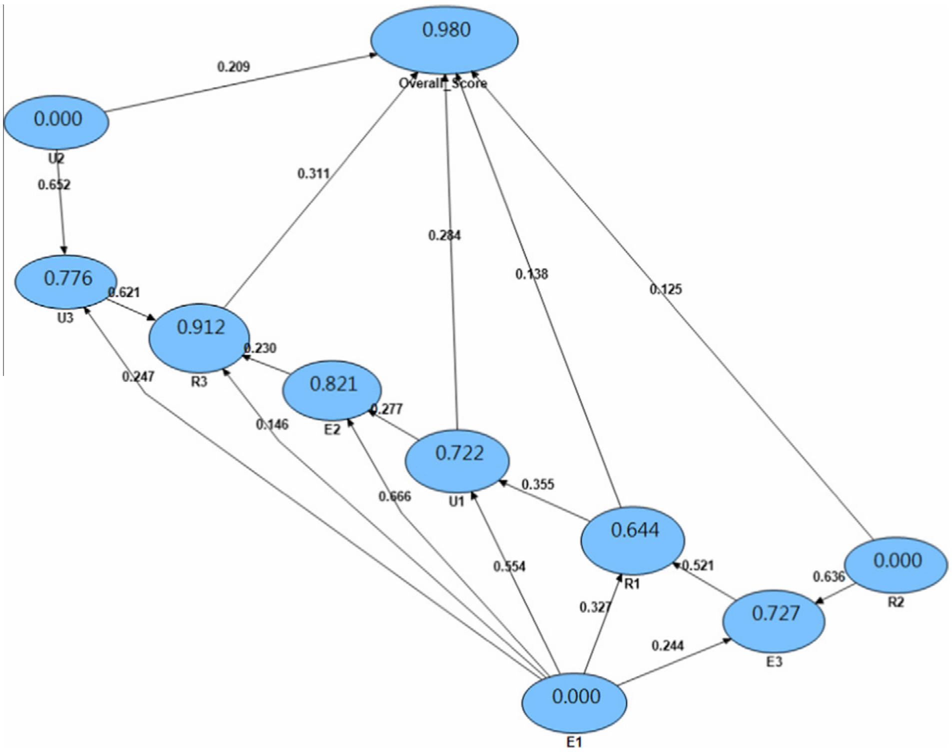

Fig. 5. The refined significant paths among pillars and Overall_Score.

global competitiveness in the medium to long term (Dutta and Mia, 2010). Our proposed approach has successfully established the casual relations among pillars and NRI overall scores of the nations' global ICT competitiveness. It has also clearly scrutinized the most critical pillars that can facilitate the policymakers to make more effective decisions. In sum, this study has contributed to add significant values to the *Annual Report*. The findings have provided informative message to the policymakers to allocate resources or effort with top priority on the most critical pillars—U2 (business usage), R2 (business readiness), E1 (market environment) and U3 (government usage) and their associated components to advance the national ICT competitiveness.

It is inevitable that this study has some limitations and calls for future studies. First, different clustering techniques would come out with different results, thus it calls for comparing the EM clustering with other proper clustering techniques in order to reach more robust conclusions. Second, the findings have discovered causal relations among nine pillars and overall score of NRI; but the conclusions are specific to the studied year. It should not be generalized or inferred to the past or in the future. Nonetheless, our study provides a good starting point to arise attentions on the issue of causal relations between pillars within NRI. In the future, it is suggested to apply the proposed approach to conducting similar analyses based on future annual reports on NRI. Third, this study does not deal with the heterogeneous samples at the observations level when using PLS path modeling, which calls for future study. Since the ICT industry has been changing drastically, it is always important to examine the consistency of the significant pillars affecting the overall NRI scores and rankings over time as well as the reality of three-stage ICT diffusion in the future studies. Furthermore, although the proposed innovative approach did not require subjectively assigning weights to the subindexes and pillars, it did not present a new set of weights for them. In the future, one can introduce appropriate methods to deal with the issue of weighing subindexes and pillars to calculate the NRI scores. Last but not least, if the components are available, one can further

Author's personal copy

Table 4
The coefficients of refined significant paths.

	Original Sample (O)	Sample Mean(M)	Standard deviation (STDEV)	Standard error (STERR)	*T* statistics (|O/STERR|)
El → E2	0.66607	0.66667	0.06258	0.06258	10.64295
El → E3	0.24433	0.24722	0.09034	0.09034	2.70450
El → R1	0.32660	0.32464	0.08488	0.08488	3.84764
El → R3	0.14587	0.14669	0.06414	0.06414	2.27420
El → U1	0.55379	0.54565	0.05714	0.05714	9.69221
El → U3	0.24689	0.25140	0.10455	0.10455	2.36132
E2 → R3	0.23040	0.23197	0.07183	0.07183	3.20745
E3 → Rl	0.52064	0.52263	0.07480	0.07480	6.96037
Rl → Overall_Score	0.13803	0.13913	0.02760	0.02760	5.00034
Rl → Ul	0.35548	0.36487	0.05578	0.05578	6.37297
R2 → E3	0.63592	0.63409	0.08733	0.08733	7.28220
R2 → Overall_Score	0.12456	0.12188	0.03403	0.03403	3.66050
R3 → Overall_Score	0.31079	0.30705	0.04211	0.04211	7.37968
Ul → E2	0.27668	0.27624	0.06691	0.06691	4.13524
Ul → Overall_Score	0.28387	0.28370	0.02186	0.02186	12.98481
U2 → Overall_Score	0.20907	0.21525	0.03575	0.03575	5.84891
U2 → U3	0.65243	0.64729	0.10475	0.10475	6.22823
U3 → R3	0.620785	0.618184	0.056449	0.056449	10.997204

develop a more complicated causal network to identify the significant components affecting the pillars. As such, the decision makers can derive critical measures (e.g., changing the value of a specific component) and make predictions with the network (e.g., seeing how the change will affect the pillars and ultimately the NRI scores). This challenging issue deserves further exploration.

Acknowledgements

The authors are indebted to three anonymous reviewers for their critical comments and constructive suggestions. Continuous encouragements and advices from the editor are also highly appreciated.

References

Anderson, R.D., Vastag, G., 2004. Causal modeling alternatives in operations research: Overview and application. European Journal of Operational Research 156 (1), 92–109.
Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., Vanthienen, J., 2004. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. European Journal of Operational Research 156 (2), 508–523.
Cerquides, J., De Mantaras, R.L., 2005. TAN classifiers based on decomposable distributions. Machine Learning 59, 323–354.
Cerulli, G., Filippetti, A., 2010. Substitutability vs. complementarity: Rethinking STI composite indicators building with an application of sensitivity analysis. Procedia – Social and Behavioral Sciences 2 (6), 7633–7635.
Chin, W.W., 1998. Issues and opinion on structural equation modeling. MIS Quarterly 22 (1).
Chow, C.K., Liu, C.N., 1968. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory 14 (3), 462–467, vii-xvi.
Curkovic, S., 2003. Environmentally responsible manufacturing: The development and validation of a measurement model. European Journal of Operational Research 146 (1), 130–155.
Dutta, S., Mia, I., 2010. The Global Information Technology Report 2009–2010. World Economic Forum and INSEAD, SRO-Kundig Geneva, Switzerland.
Fernandez, E., Navarro, J., Bernal, S., 2010. Handling multicriteria preferences in cluster analysis. European Journal of Operational Research 202 (3), 819–827.
Fornell, C., Bookstein, F., 1982. Two structural equations models: LISREL and PLS applied to consumer exit-voice theory. Journal of Marketing Research 19 (4), 440–452.
Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Machine Learning 29 (2–3), 131–163.
Grupp, H., Mogee, M.E., 2004. Indicators for national science and technology policy: How robust are composite indicators? Research Policy 33 (9), 1373–1384.
Grupp, H., Schubert, T., 2010. Review and new evidence on composite innovation indicators for evaluating national performance. Research Policy 39 (1), 67–78.
Henseler, J., Ringle, C.M., Sinkovics, R.R., 2009. The use of partial least squares path modeling in international marketing. Advances in International Marketing 20, 277–319.
Hruschka, E.R., Ebecken, N.F.F., 2007. Towards efficient variables ordering for Bayesian networks classifier. Data & Knowledge Engineering 63 (2), 258–269.
Hulland, J., 1999. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. Strategic Management Journal 20 (2), 195–204.
Jakobowicz, E., Derquenne, C., 2007. A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy. Computational Statistics & Data Analysis 51 (8), 3666–3678.
Karaboga, D., Ozturk, C., 2011. A novel clustering approach: Artificial bee colony (ABC) algorithm. Applied Soft Computing 11 (1), 652–657.
Lauria, E.J.M., Duchessi, P.J., 2007. A methodology for developing Bayesian networks: An application to information technology (IT) implementation. European Journal of Operational Research 179 (1), 234–252.
Lee, S.Y.T., Kim, H.W., Gupta, S., 2009. Measuring open source software success. Omega 37 (2), 426–438.
Nadkarni, S., Shenoy, H.P., 2001. A Bayesian network approach to making inferences in causal maps. European Journal of Operational Research 128 (3), 479–498.
Nascimento, M.C.V., de Carvalho, A.C.P.L.F., 2011. Spectral methods for graph clustering: A survey. European Journal of Operational Research 211 (2), 221–231.
Ranganathan, C., Sethi, V., 2002. Rationality in strategic information technology decisions: The impact of shared domain knowledge and IT unit structure. Decision Sciences 33 (1), 59–86.
Sohn, S.Y., Han, H.K., Jeon, H.J., 2007. Development of an air force warehouse logistics index to continuously improve logistics capabilities. European Journal of Operational Research 183 (1), 148–161.
Su, Y.F., Yang, C., 2010. A structural equation model for analyzing the impact of ERP on SCM. Expert Systems with Applications 37 (1), 456–469.
Temme, D., Kreis, H., Hildebrandt, L., 2006. PLS path modeling: A software review. SFB 649 Discussion Papers SFB649DP2006-084, Humboldt University, Berlin, Germany.
Tenenhaus, M., Vinzi, V.E., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. Computational Statistics and Data Analysis 48 (1), 159–205.
Verma, T.S., Pearl, J., 1991. Equivalence and synthesis of causal models. In: Bonissone, P., Henrion, M., Kanal, L., Lemmer, J.F. (Eds.), Uncertainty in Artificial Intelligence, vol. 6. Elsevier, Berlin, pp. 255–268.
Wang, T., Touchman, J.W., Xue, G., 2004. Applying two-level simulated annealing on Bayesian structure learning to infer genetic networks. In: Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference, 2004. pp. 647–648.
Witten, I., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufmann, San Francisco.
Wixom, B.H., Watson, H.J., 2001. An empirical investigation of the factors affecting data warehousing success. MIS Quarterly 25 (1), 17–41.
Wu, W.W., 2010. Linking Bayesian networks and PLS path modeling for causal analysis. Expert Systems with Applications 37 (1), 134–139.
Wu, W.W., 2011. Improving classification accuracy and causal knowledge for better credit decisions. International Journal of Neural Systems 21 (4), 297–309.