



Evolutionary attribute ordering in Bayesian networks for predicting the metabolic syndrome

Han-Saem Park, Sung-Bae Cho*

Department of Computer Science, Yonsei University, 262 Seongsanno, Sudaemoon-ku, Seoul 120-749, South Korea

ARTICLE INFO

Keywords:

Bayesian network
Prognostic modeling
Attribute ordering optimization
Metabolic syndrome

ABSTRACT

The metabolic syndrome is a set of risk factors that include abdominal obesity, insulin resistance, dyslipidemia and hypertension. It has affected around 25% of adults in the US and become a serious problem in Asian countries recently due to the change in dietary habit and life style. On the other hand, Bayesian networks that are the models to solve the problems of uncertainty provide a robust and transparent formalism for probabilistic modeling, so they have been used as a method for diagnostic or prognostic model in medical domain. Since the K2 algorithm, a well-known algorithm for Bayesian networks structure learning, is influenced by an input order of the attributes, an optimization of BN attribute ordering has been studied as a research issue. This paper proposes a novel ordering optimization method using a genetic algorithm based on medical expert knowledge in order to solve this problem. For experiments, we use the dataset examined twice in 1993 and 1995 in Yonchon County of Korea. It has 18 attributes of 1193 subjects participated in both surveys. Using this dataset, we make the prognostic model of the metabolic syndrome using Bayesian networks with an optimized ordering by evolutionary approach. Through an ordering optimization, the prognostic model of higher performance is constructed, and the optimized Bayesian network model by the proposed method outperforms the conventional BN model as well as neural networks and k -nearest neighbors. Finally, we present the application program using the prognostic model of the metabolic syndrome in order to show the usefulness of the proposed method.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The metabolic syndrome is composed of a cluster of metabolic disorders including abdominal obesity, insulin resistance, dyslipidemia and hypertension, and the correlation between metabolic syndrome and other diseases such as diabetes and coronary heart disease is reported in the literature (Cabre et al., 2008; Lee et al., 2009). It affects around 25% of adults over the age of 20 and up to 45% over age 50 in the United States (Mehta & Reilly, 2004). These days, it is found even in children and adolescents as a ratio of approximately 4%–7% (Pan & Pratt, 2008). In Asian countries, it has become a significant problem lately due to the change in dietary habit and life style (Moon et al., 2003; Son, Kunii, Hung, Sakai, & Yamamoto, 2005). In situations like this, many groups have been studying the metabolic syndrome from all over the world (Cabre et al., 2008; Lee et al., 2009; Mehta & Reilly, 2004; Moon et al., 2003; Pan & Pratt, 2008; Son et al., 2005).

Recently, computer based health-care systems have been studied and developed a lot according to an advancement of artificial intelligence techniques such as image processing and expert sys-

tems and increase of people's concern for their health (Herzbert et al., 2009). The systems for computer-based diagnosis or pervasive health-care are the representative examples (Innocent & John, 2004; Osmani, Balasubramaniam, & Botvich, 2008). The Bayesian network, one of the AI techniques, has emerged in recent years as a powerful technique for handling uncertainty in complex domains (Larranaga, Poza, Yurranmendi, Murga, & Kuijpers, 1996). It is a model of a joint probability distribution over a set of random variables. The Bayesian network is represented as a directed acyclic graph where nodes correspond to variables and arcs correspond to probabilistic dependencies between connected nodes (Chen & Blanchette, 2007). Bayesian networks have been used for prediction or classification problem in the medical domain and shown high performance. In particular, they have been applied successfully to the modeling of diagnosis and prognosis for diverse diseases (Antal, Fannes, Timmerman, Moreau, & Moor, 2004, 2003; Aronsky & Haug, 2000; Charitos, Gaag, Visscher, Schurink, & Lucas, 2009; Gerven, Jurgelenaite, Taal, Heskes, & Lucas, 2007; Getoor, Rhee, Koller, & Small, 2004; Maskery, Hu, Hooke, Shriver, & Liebman, 2008; Sierra et al., 2001; Tucker, Vinciotti, Liu, & Garway-Heath, 2005; Wang, Zheng, Good, King, & Chang, 1999). There have been many black box tools that classify or predict several diseases, and neural networks are the representative example. Bayesian networks have

* Corresponding author. Tel.: +82 2 2123 2720; fax: +82 2 365 2579.

E-mail address: sbcho@cs.yonsei.ac.kr (S.-B. Cho).

strengths that they can use the domain knowledge easily and analyze the results compared to them (Wang et al., 1999). Even though they are sometimes not better than neural networks in terms of accuracies, Bayesian networks are appropriate methods in the medical domain that requires to analyze the results with medical knowledge.

This paper used a genetic algorithm to solve an optimization problem in medical prognostic modeling. Specifically, this paper deals with a problem that predicts the metabolic syndrome with the dataset obtained in Yonchon County of Korea. This paper constructs a prognostic model using Bayesian network, and has used the K2 algorithm by Cooper and Herskovits in order to learn its structure (Larranaga, Poza, et al., 1996). Since the result of the K2 algorithm is influenced by an input ordering of the attributes, an optimization of this ordering has been also studied (Hruschka & Ebecken, 2007; Hruschka, Santos, & Galvao, 2007; Hsu, 2004; Larranaga, Kuijpers, Murga, & Yurramendi, 1996; Song, Eder, & Nguyen, 2007). This paper proposes an efficient optimization method using medical domain knowledge and a genetic algorithm in order to solve this problem. Different from the conventional methods, after clustering similar attributes into each group, an ordering of the groups and an ordering of the attributes in each group have been performed in turns. As applying the medical domain knowledge, an efficient and reliable modeling has been conducted. Subsequently, the experiments using the proposed prognostic model have been conducted after the structure and parameter learning processes, and an application program using this model has been presented in order to show its usefulness.

This paper is organized as follows. Section 2 presents the definition of the metabolic syndrome, the advantages of BN application to medical domains and related works on Bayesian network attribute ordering. In Section 3, the proposed method using medical domain knowledge and the genetic algorithm is described. In Section 4, various experimental results in order to show the usefulness of the proposed method are provided, and Section 5 concludes the paper with summary.

2. Backgrounds

2.1. Metabolic syndrome

The definition of the metabolic syndrome was provided by the National Cholesterol and Education Program, Adult Treatment Panel III (ATP III). It requires the presence of three or more of the following components (Son et al., 2005):

- (1) Abdominal obesity (waist circumference >102 cm in men and >88 cm in women).
- (2) Hypertriglyceridemia (≥ 150 mg/dL).
- (3) Low high-density lipoprotein (HDL) cholesterol (<40 mg/dL in men and <50 mg/dL in women).
- (4) High blood pressure (systolic ≥ 130 mm Hg or diastolic ≥ 80 mm Hg).
- (5) High fasting glucose (>110 mg/dL).

Because this original standard is not appropriate for Asian, we use a modified definition for Asian of the abdominal obesity (waist circumference >90 cm in men and >80 cm in women) (Moon, Cho, Lim, Park, & Lee, 2003) in this paper.

2.2. Bayesian networks in medical domain

Bayesian networks have several advantages as a data mining tool. First, they allow researchers to use their domain knowledge in the discovery process, while other techniques mainly rely on

coded data to extract knowledge. Second, the models made by Bayesian networks can be more easily understood than most of other models via the use of nodes and arches. Third, they are superior in capturing interactions among input variables. Bayesian networks are so flexible that they can make models even though there are a few missing values in the data. In addition, Bayesian networks are less influenced by small sample size since they can incorporate domain knowledge into statistical data (Lee & Abbott, 2003).

Based on advantages mentioned above, Bayesian networks have been utilized for diagnosis and prediction of diseases in medical domain. Antal et al. used the Bayesian network to construct diagnostic model of ovarian cancer and to classify its samples (Antal et al., 2004; Antal, Fannes, Timmerman, Moreau, & Moor, 2003). Bayesian networks were also applied to diagnosis of pneumonia (Aronsky & Haug, 2000; Charitos et al., 2009) and breast cancer (Maskery et al., 2008; Wang et al., 1999). Also, Bayesian network has used to model the relationships between attributes. Tucker et al. made use of large amount of clinical data in order to build models with spatio-temporal nature of those data (Tucker et al., 2005), and Gerven et al. employed a novel Bayesian model that facilitates the interpretation of cause and effect relationships (Gerven et al., 2007). Besides, studies for several purposes such as patient caring, tuberculosis model, and Getoor et al. and Sierra et al. have used Bayesian networks to solve the problems in the medical domain (Getoor et al., 2004; Sierra et al., 2001).

2.3. Attribute ordering in Bayesian networks

This paper uses the K2 algorithm for learning of Bayesian network model, and the performance of this algorithm is influenced by ordering of input attributes. Searching for a good attribute ordering can be conducted by (1) expert knowledge, (2) exhaustive search and (3) heuristic search like a genetic algorithm (Hsu, 2004). Most works have focused only on the last approach for attribute ordering in Bayesian networks since Larranaga et al. presented one (Hruschka & Ebecken, 2007; Hruschka et al., 2007; Hsu, 2004; Larranaga, Kuijpers, et al., 1996). For example, Hsu presented genetic wrappers for selection and ordering of attributes (Hsu, 2004). For attribute ordering, he implemented a genetic algorithm for the permutation problem of attribute ordering for Bayesian network structure learning using the K2 algorithm. Hruschka et al. proposed wrapper methods using a genetic algorithm and Markov blanket (Hruschka et al., 2007). It reduced the time cost of GA-based attribute ordering method as using MarkovPC algorithm, a constrained PC algorithm using Markov blanket (Song et al., 2007). Not all works used wrapper approach. Hruschka also proposed a simple but efficient attribute ordering method for Bayesian networks based on feature ranking algorithm such as chi-square and information gain (Hruschka & Ebecken, 2007). However, it cannot be more accurate than wrapper methods, which are more general for attribute ordering in Bayesian networks. In this paper, we do not only use a genetic algorithm for attribute ordering in Bayesian networks but domain expert knowledge for performance and efficiency. It enhances the strength of the wrapper method as making ordering process be more efficient and reliable using domain knowledge.

3. Proposed method

A basic idea of the proposed method is based on our previous work (Park & Cho, 2006). This paper refines the previous method with more sophisticated process, better justification and more in-depth analyses of the experimental results. An application program using the proposed method is also developed. This section describes

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات