



Diagnose the mild cognitive impairment by constructing Bayesian network with missing data

Yan Sun^{a,b,c}, Yiyuan Tang^{a,*}, Shuxue Ding^d, Shipin Lv^c, Yifen Cui^a

^a Neuroinformatics Institute, Dalian University of Technology, Dalian 116024, China

^b Department of Computer Science, Liaoning Normal University, Dalian 116029, China

^c Department of Computer Science and Engineering, Dalian University of Technology, Dalian 116024, China

^d School of Computer Science and Engineering, Aizu University, Aizu-Wakamatsu City, Fukushima 965-8580, Japan

ARTICLE INFO

Keywords:

Mild cognitive impairment (MCI)
Missing data
Bayesian network
Mutual information
Newton interpolation

ABSTRACT

Mild Cognitive Impairment (MCI) is thought to be the prodromal phase to Alzheimer's disease (AD), which is the most common form of dementia and leads to irreversible neurogenerative damage of the brain. In order to further improve the diagnostic quality of the MCI, we developed a MCI expert system to address MCI's prediction and inference question, consequently, assist the diagnosis of doctor. In this system, we mainly deal with following problems: (1) Estimate missing data in the experiment by utilizing mutual information and Newton interpolation. (2) Make certain the prior feature ordering in constructing Bayesian network. (3) Construct the Bayesian network (We term the algorithm as MNBN). The experimental results indicate that MNBN algorithm achieved better results than some existing methods in most instances. The mean square error comes to 0.0173 in the MCI experiment. Our results shed light on the potential application in MCI diagnosis.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia and that may lead to irreversible neurogenerative damage of the brain. But the current diagnostic tools have poor sensitivity, especially for the early stages of AD and are not easy to be diagnosed until AD has led to irreversible brain damage (Morris et al., 2001). Therefore, it is very important research topic for how to diagnose AD as early as possible. Through research effort of recent 10 years, it is concluded that MCI (Mild cognitive impairment) is the early stage of the Alzheimer Diseases (Celsis, 2000; Morris et al., 2001; Petersen et al., 2001). 10–30% of MCI patients convert to AD annually, whereas the rate of conversion of cognitively normal elderly people is 1–2% (Celsis, 2000). Furthermore, there is evidence that 100% of patients with MCI progress to greater dementia severity (Petersen et al., 2001). So the problem of diagnosing AD can be converted into the diagnosis of the MCI. Up to now, however, there is still not a strict and unified standard.

In this study, we develop a specific diagnostic system on the MCI, which predicts and diagnoses the MCI by using some artificial intelligent methods. Since a practical database usually might be not complete, at first, we utilize the mutual information and

Newton interpolation to estimate the values of missing data. Then, we propose to determine the feature ordering by using the mutual information and defining a “higher filter”. Finally, we construct the Bayesian network for assisting the prediction and diagnosis of the MCI.

The remainder of this paper is organized as follows: Section 2 briefly reviews some related works. In Section 3, we present the MNBN algorithm. In Section 4, we further describe how to implement the MNBN algorithm. In Section 5, we report and analyze experimental results. Finally, in Section 6, we draw the main conclusions and give some discussions.

2. Related works

In recent years, one new idea is to assistant diagnose the MCI by using some method of artificial intelligence. Among them, Bayesian Network is popular within the community of artificial intelligence due to their ability to support probabilistic reasoning from data with uncertainty. According to the network, probabilistic inference can be conducted to predict the values of some variables based on the observed values of other variables. Hence, Bayesian networks are widely used in many areas. Reference (Chen & Herskovits, 2006) applied Bayesian Network to model the interactions among morphological changes and clinical variables of the MCI. We can conclude four principal advantages of using a discrete variable Bayesian network for network analysis:

* Corresponding author. Address: Neuroinformatics Institute, Dalian University of Technology, Dalian, No. 2 Linggong Road, Ganjingzi District, Dalian City, Liaoning Province 116024, China. Tel.: +86 411 84706039; fax: +86 411 84706046.

E-mail address: yiyuan@uoregon.edu (Y. Tang).

(1) the Bayesian network framework does not require that the joint distribution follows a specific parametric distribution; (2) a Bayesian network supports probabilistic reasoning as it consists of probabilistic associations among variables; (3) because the Bayesian network representation is based on the concept of conditional independence, it supports Bayesian inference without having to maintain the full joint distribution in memory; and (4) since each Bayesian network is a multivariate model that we can evaluate using a single probability score, we can evaluate many structure–function interactions without the multiple comparison problem.

Although the method of network analysis is very effective, such as the method of reference (Wong & Leung, 2004; Chen & Herskovits, 2006; Liang & Zhang, 2009), these algorithms can not deal with the missing data. Whereas, in real-world applications, missing data is a great quantity, especially in the medical problem. In the MCI experiment, because the eyesight of some subjects is not better, they can not see clearly the stimulation, which leads to appear some missing values in the data set. One example of the importance of handling missing data is that more than 40% of data sets in the UCI repository have missing values (Newman, Hettich, & Blake, 1998; Garcia-Laencina, Sancho-Gomez, Figueiras-Vidal, & Verleysen, 2009), which is one of most commonly used data sets for benchmarking machine learning procedures. Many high efficient and effective learning algorithms require complete data sets to execute. The conversion from an incomplete data set to a complete one then becomes an issue.

Many researchers have been working on constructing Bayesian network from incomplete data sets. However, there are a few algorithms available for learning the Bayesian network structure with missing data, because most algorithms require a complete data set (Lin & Haug, 2008). For learning Bayesian network from incomplete data set, the most important challenge is that the parameter values and the scores of networks can not be computed directly on the cases with missing values. Moreover, the scoring metric can not be decomposed directly. Thus, a local change in the network structure will lead to reevaluate the score of the whole network (Wong & Guo, 2008).

Friedman proposes a Bayesian Structural Expectation Maximization (SEM) algorithm which alternates between the parameter optimization process and the model search process (Friedman, 1997, 1998). However the algorithm takes much longer time to run and is lack of stability (Austin & Escobar, 2005; Lin and Haug, 2008). The score of a Bayesian network is maximized by means of the maximization of the expected score. Pena et al. uses the BC + EM method instead of the EM method (Dempster, Laird, & Rubin, 1977) in their BS-BC + EM algorithm for clustering (Pena, Lozano, & Larranaga, 2000; Pena, Lozano, & Larranaga, 2002). However, the search strategies adopted in most existing SEM algorithms may not be effective and may make the algorithms find sub-optimal solutions. Myers et al. employ a genetic algorithm to learn Bayesian networks from incomplete data sets (Myers, Laskey, & DeJong, 1999). Both network structures and the missing values are encoded and evolved. The incomplete data set is completed by specific genetic operators during evolution. Nevertheless, it has the efficiency and convergence problems because of the enlarged search space and the strong randomness of the genetic operators for completing the missing values.

It is worth mentioning that Wong uses evolutionary algorithm to learn Bayesian network from incomplete data sets, called EBN (Evolutionary Bayesian Network learning method) (Wong & Guo, 2008), which utilizes the efficient and effective global search ability of HEA (Wong & Leung, 2004) and applies EM (Dempster et al., 1977) to handle missing values. However, EBN is a stochastic algorithm and results are strongly dependent on the initial network structure, so the results are not stationary.

In this paper, we propose a novel method that firstly uses mutual information to get the important extent of the feature. According to the importance of feature we find the most similar cases with the missing value case. Then, we adopt the Newton interpolation to estimate the value of the missing data. Finally, we construct the Bayesian network by using K2 algorithm (Cooper & Herskovits, 1992), it is the most effective, efficient and most popular. However, this algorithm has one disadvantage that it must specify a prior feature sequence. The feature ordering consists of domain knowledge or constraints that specify a partial order, such that a parent feature must appear earlier in the order than any of its descendants (Chen & Herskovits, 2006). The prior feature ordering most depends on the subjective experience of researchers, which serious effect on the results of the Bayesian network model. Some researcher used the oriented tree obtained from the maximum-weight spanning-tree algorithm to generate this ordering (Heckerman, Geiger, & Chickering, 1994; Chen & Herskovits, 2006). In this study, we utilize the mutual information and define a “higher filter” to learn the prior feature ordering.

3. The algorithm

We suppose that there is a data set D (or sampling space) with $X = \{x_1, \dots, x_n\} \subset R^d$, for each case $x_a \in X (a = 1, \dots, n)$ has m features $F = \{f_1, \dots, f_m\}$, it can be represented as a value vector of features, i.e., $x_a = (v_{a1}, \dots, v_{am})$, where v_{ai} is the value of x_a corresponding to the feature f_i . Among them some v_{ai} are missing and the number of missing data is k .

Given a data set D , the first objective of learning algorithm is to get the estimation of the missing data v_{ai} by computing the mutual information and Newton interpolation. The second objective is to get the prior feature ordering by defining a “higher filter”. The third goal is to construct the Bayesian network B_s using K2 algorithm (Cooper & Herskovits, 1992), it will find the nonlinear relationships among all the features and get the posterior distribution for the functional feature MCI from the Bayesian network. That is, we can predict the state of feature MCI based on the Bayesian network.

3.1. Getting the relationships between features

Since mutual information is good at quantifying how much information is shared by two random variables, it is often taken as evaluation criterion to measure the relevance between features and the class labels (Marcus, Hutter, & Zaffalon, 2005; Liu, Sun, Liu, & Zhang, 2009). In this study, we utilize mutual information to measure the relationship between features, which aims to estimate the missing feature data.

We assume that $f_i \subseteq F$ and $f_j \subseteq F (1 \leq i, j \leq m)$ represent the selected missing value feature and candidate feature subsets, respectively. According to the definition of the mutual information, we will get the mutual information between f_i and f_j by Eq. (1).

$$I(f_i; f_j) = \sum_{f_i \in F} \sum_{f_j \in F, f_j \neq f_i} p(f_i, f_j) \log \frac{p(f_i, f_j)}{p(f_i)p(f_j)}. \quad (1)$$

Under this context, those features $f_j \in F$ with high predictive power will have larger mutual information $I(f_i; f_j)$. On the contrary, $I(f_i; f_j)$ is zero if f_i and f_j are independent with each other. At this point, f_j has no contribution to the distribution of f_i .

3.2. Finding the most similar cases

In this section, we use the mutual information $I(f_i; f_j)$ as weights to find the most similar cases with the missing values case x_a .

Firstly, in order to discard the feature that is irrelevant or weakly relevant with the selected missing value feature f_i , we first

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