Financial Data Modeling using a Hybrid Bayesian Network Structured Learning Algorithm

Shun Li, Peking University, China

Da Shi, Peking University, China

Shaohua Tan, Peking University, China

ABSTRACT

In this paper, a group of hybrid incremental learning algorithms for Bayesian network structures are proposed. The central idea of these hybrid algorithms is to use the polynomial-time constraint-based technique to build a candidate parent set for each domain variable, followed by the hill climbing search procedure to refine the current network structure under the guidance of the candidate parent sets. Experimental results show that, the authors' hybrid incremental algorithms offer considerable computational complexity savings while obtaining better model accuracy compared to the existing incremental algorithms. One of their hybrid algorithms is also used to model financial data generated from American stock exchange markets. It finds out the predictors of the stock return among hundreds of financial variables and, at the same time, the authors' algorithm also can recover the movement trend of the stock return.

Keywords: *Bayesian Networks, Incremental Learning, Sequential Prediction, Stock Return Prediction, Structure Learning*

INTRODUCTION

Bayesian Network and other Bayesian models are becoming increasingly prominent across a broad spectrum of the cognitive sciences (Perfors, Tenenbaum, Griffiths, & Xu, 2011) and cognitive informatics (Wang, 2003, 2007, 2009, 2011; Wang & Kinsner, 2006; Wang, Kinsner, & Zhang, 2009; Wang et al., 2009). In recent years, these Bayesian models have been used to address animal learning (Courville, Daw, & Touretzky, 2006), human inductive learning and generalization (Tenenbaum, Griffiths, & Kemp, 2006), visual scene perception (Yuille & Kersten, 2006), motor control (Kording & Wolpert, 2006), semantic memory (Steyvers, Griffiths, & Dennis, 2006), language processing and acquisition (Chater & Manning, 2006), symbolic reasoning (Oaksford & Chater, 2001; Wang, 2011), causal learning and inference (Steyvers, Tenenbaum, Wagenmakers,

DOI: 10.4018/jcini.2012010103

& Blum, 2003; Griffiths & Tenenbaum, 2005; Griffiths & Tenenbaum, 2007; Pacer & Griffiths, 2011; Buchsbaum, Gopnik Griffiths & Shafto, 2011), social cognition (Baker, Tenenbaum, & Saxe, 2007), and some other cognitive problems. However, all the Bayesian network models used in these researches are static models, which only can represent an average situation during special time. This paper focuses on incremental Bayesian network structure learning algorithms, which can be used to dynamically describe the changes of the cognitive process in time.

Over a decade of research in finding efficient incremental learning algorithms for Bayesian network structures has yielded quite a number of important results and computational algorithms (Buntine, 1991; Friedman & Goldszmidt, 1997; Lam, 1998; Lam & Bacchus, 1994b; Nielsen & Nielsen, 2007; Roure, 2004a, 2004b; Shi & Tan, 2007; Niinimaki, Parviaimen, & Koivisto, 2011; Pennock & Xia, 2011). It is well recognized, however, that these existing algorithms often suffer from high computational complexity which prevents them from solving complex and large-scale practical problems.

In this paper, a hybrid learning template is proposed to overcome the computational complexity deficiencies, and a group of algorithms are developed based on the template. Our template consists of polynomial-time constraint-based techniques and the hill climbing search procedure to decouple the complexity into two smaller and less complex computations. The constraint-based techniques make best use of the information in the current network structure and newly arrived datasets to build compact candidate parent sets for domain variables. The hill climbing search procedure is then employed to refine the current network structure under the guidance of the candidate parent sets.

Two kinds of constraint-based techniques are proposed. The first one builds candidate parent sets from a global view. It learns an undirected tree-shaped network structure, and then extracts candidate parent sets based on the undirected structure and the current network structure simultaneously. The second one builds candidate parent sets from a local view. It selects the most relevant variables to each domain variable as the candidate parents by using polynomial-time feature selection algorithms. Two kinds of hybrid incremental algorithms are developed based on the above two constraint-based techniques.

An extensive comparative study of our algorithms, both analytically and computationally, against a wide cross-section of state-of-the-art incremental algorithms and some batch algorithms on classic benchmark datasets is carried out. To our knowledge, this is the first comprehensive study providing a valuable comparison of the existing algorithms for various datasets with different sample size. The results show that our hybrid algorithms, especially the one based on the global view, offer considerable computational complexity savings compared to the existing algorithms all the time, and some of them also obtain better model accuracy at the same time.

To further inspect the computational complexity and the validation of our algorithm, the one based on the global view is also used to solve real-world stock return prediction problem. The stock return prediction, including both the future stock return value prediction and the future stock return movement trends prediction, has gained unprecedented popularity in financial market forecasting research in recent years (Avramov & Chorida, 2006a, 2006b; Banz, 1980; Basu, 1977; Fama & French, 1989, 1992; Jegadeesh, 1990; Jegadeesh & Titman, 1993; Keim & Stambaugh, 1986; Lettau & Ludvigson, 2001). Finding out predictors of the stock return exactly is the foundation of the above two tasks. Unfortunately, there are no good solutions for this problem till now. Our algorithm uses Bayesian network structures to search the predictors automatically, which provides a brand new way of solving this prediction problem.

The rest of this paper is organized as follows: first, our hybrid incremental algorithms are introduced. The experiments are described afterwards, followed by a simple theory analysis.

Related Content

Development of an Ontology for an Industrial Domain
Christine W. Chan (2007). *International Journal of Cognitive Informatics and Natural Intelligence (pp. 36-51).*
www.irma-international.org/article/development-ontology-industrial-domain/1539

Cognitive Processes by using Finite State Machines
Ismael Rodríguez, Manuel Núñezand Fernando Rubio (2009). *Novel Approaches in Cognitive Informatics and Natural Intelligence (pp. 52-64).*
www.irma-international.org/chapter/cognitive-processes-using-finite-state/27298

Cognitive Dynamic Systems
Simon Haykin (2011). *International Journal of Cognitive Informatics and Natural Intelligence (pp. 33-43).*
www.irma-international.org/article/cognitive-dynamic-systems/63620

Structure of the Relational Thinking Styles Model
(2012). *Relational Thinking Styles and Natural Intelligence: Assessing Inference Patterns for Computational Modeling (pp. 43-61).*
www.irma-international.org/chapter/structure-relational-thinking-styles-model/65041

A Cognitive Approach to the Mechanism of Intelligence
Yi X. Zhong (2008). *International Journal of Cognitive Informatics and Natural Intelligence (pp. 1-16).*
www.irma-international.org/article/cognitive-approach-mechanism-intelligence/1550