



Knowledge-based data mining of news information on the Internet using cognitive maps and neural networks

Taeho Hong*, Ingoo Han

Korea Advanced Institute of Science and Technology, Graduate School of Management, 207-43 Cheongryangri-Dong, Dongdaemun-Gu, Seoul 130-012, South Korea

Abstract

In this paper, we investigate ways to apply news information on the Internet to the prediction of interest rates. We developed the Knowledge-Based News Miner (KBNMiner), which is designed to represent the knowledge of interest rate experts with cognitive maps (CMs), to search and retrieve news information on the Internet according to prior knowledge, and to apply the information, which is retrieved from news information, to a neural network model for the prediction of interest rates.

This paper focuses on improving the performance of data mining by using prior knowledge. Real-world interest rate prediction data is used to illustrate the performance of the KBNMiner. Our integrated approach, which utilizes CMs and neural networks, has been shown to be effective in experiments. While the 10-fold cross validation is used to test our research model, the experimental results of the paired *t*-test have been found to be statistically significant. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Data mining; Internet; Cognitive maps; Neural networks

1. Introduction

Nowadays, the capability to both generate and collect data has been expanded enormously and provides us with huge amounts of data. Millions of databases are being used in business data management, scientific and engineering data management, as well as other applications. Data mining has become a research area with increasing importance with the amount of data greatly increasing (Changchien & Lu, 2001; Chiang, Chow, & Wang, 2000; Fayyads, Piatetsky-Shapiro, & Smyth, 1996; Park, Piramuthu, & Shaw, 2001). Furthermore, data mining has come to play an important role since research has come to improve many methods used in data mining applications including statistical pattern recognition, association rules, recognizing sequential or temporal patterns, clustering or segmentation, data visualization, and classification.

Although most data is stored in a database from which it can readily be applied to a data mining application, some kinds of data such as news information is not. As the popularity of the World Wide Web increases, many newspapers expand their services by providing news information on the web in order to be more competitive and increase benefits. The web disseminates real time news to investors. News

information includes articles on the political situation, social conditions, international events, government policies, trader's psychology, and all those topics, which we see and understand through the Internet. Such information is formulated in the form of texts, referred to as documents, and thus text mining is required if the information is to be applied in data mining applications.

Many researchers attempt to predict interest rates by using the time series model (Bidarkota, 1998), neural networks model (Hong & Han, 1996), the integrated model of neural networks and case-based reasoning (Kim & Noh, 1997). Meanwhile another approach was attempted in the prediction of the stock price index where Kohara, Ishikawa, Fukuhara, and Nakamura (1997) took into account non-numerical factors such as political and international events from newspaper information. They insist that, with event information acquired from newspapers, this method improves prediction ability of neural network. Although they personally read newspapers and rated each political and international event according to their judgment, it is, however, not easy for people to search and retrieve the vast amount of news simply through his/her knowledge and capacity. So we propose a means of applying news information from the Internet for the prediction of interest rates. The system discussed here, named the Knowledge-Based News Miner (KBNMiner), is designed to adopt a prior knowledge base, representing expert

* Corresponding author. Tel.: +82-2-958-3131; fax: +82-2-958-3604.
E-mail addresses: hongth@kgs.m.kaist.ac.kr (T. Hong).

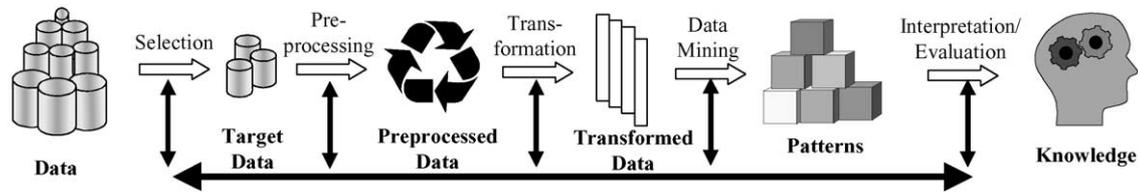


Fig. 1. Overview of the steps constructing the KDD process 0.

knowledge, as a foundation on which to probe and collect news and then to apply this news information to a neural network model for interest rate predictions.

A cognitive map (CM) is used to build the prior knowledge base. CM is a representation perceived to exist by a human being in a visible or conceptual target world. CM manages the causality and relation of non-numeric factors mentioned earlier. The KBNMiner retrieves the event information from news information on the web utilizing CM and prior knowledge. Event information is divided into two types in the KBNMiner. One is positive event information, which affects the increase of interest rates, and the other is negative event information, which affects the decrease of interest rates. A neural network model is developed and experimented on using event information.

This study focuses on the effect news information can have on the prediction of interest rates. As discussed earlier, the event information, which is acquired by the KBNMiner, is applied into a neural network model for the validation of our suggested method. More, specifically, the following research question is addressed:

- What is the effect of the event information on the neural network performance when compared to other prediction models with no event information such as the neural network and random walk models?

In Section 2, we provide a brief overview of data mining and discuss the CM method employed in KBNMiner and the way to build prior knowledge with CMs. Section 3 introduces the architecture of KBNMiner and presents a detailed description of KBNMiner. In Section 4, interest rate prediction data is used to illustrate the performance of KBNMiner. And we present the results of our approach and analyze the results statistically. Finally, the conclusion is presented.

2. Data mining and knowledge engineering

2.1. Data mining

Data mining has become a research area with increasing importance (Changchien & Lu, 2001; Chiang et al., 2000; Fayyads et al., 1996; Park et al., 2001). Berry and Linoff (1997) defines data mining as the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules. Frawley, Piatetsky-shapiro, and Matheus (1991) refer to

the entire process involving data mining as knowledge discovery in database (KDD). They view the term data mining as referring to a single step in the process that involves finding patterns in the data. However, Allen (1996) notes ‘data mining is the entire process of knowledge discovery’. Fayyads et al. (1996) outline a practical view of the data mining process emphasizing its interactive and iterative nature in Fig. 1. The KDD process is summarized as: (1) Learning the application domain, (2) Creating a target data set, (3) Data cleaning and preprocessing, (4) Data reduction and projection, (5) Choosing the function of data mining, (6) Choosing the data mining algorithms, (7) Data mining, (8) Interpretation, (9) Using discovery knowledge. The core of the knowledge discovery process is the set of data mining tasks used to extract and verify patterns in data. However, this core typically composes only a small part (estimated at 15–25%) of the effort of the overall process (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro, & Simoudis, 1996).

Specific techniques used in data mining applications include market basket analysis, memory based reasoning, cluster detection, link analysis, decision trees and rule induction, neural networks, and genetic algorithms, etc. However, most of existing algorithms are primarily data-driven and do not fully exploit domain knowledge and intuition that decision makers in business environment have (Padmanabhan & Tuzhilin, 1999). Data mining with prior knowledge is expected to exhibit superior performance than data mining without. This suggests a need for methods to initially adopt a prior knowledge base in data mining applications and this study thus develops a framework and the KBNMiner system to address this issue.

2.2. Knowledge engineering and cognitive maps

Knowledge is an interesting concept that has attracted the attention of philosophers for thousands of years. In more recent times, researchers have investigated knowledge in a more applied way with the chief aim of bringing knowledge to life in machines. Artificial intelligence has contributed to the perceived challenge by developing new tools to produce knowledge from data. However, knowledge is a complex concept and is itself, invisible. These two factors lead to difficulties in the attempt to manage knowledge. One of the more serious problems is that knowledge is built differently among human beings corresponding to their common experiences. People have knowledge consisting

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