Chapter XV

Data Mining Using Qualitative Information on the Web

Taeho Hong
Pusan National University, Korea

Woojong Suh
Inha University, Korea

Abstract

Data mining has drawn much attention in generating the useful information from Web data. Data mining techniques have typically considered quantitative information rather than qualitative, though the qualitative information can often be used to improve the quality of a result. This chapter provides a hybrid data mining application, KBNMiner (Knowledge-Based News Miner), to predict interest rates on the basis of qualitative information on the Web as well as quantitative information stored in a database. The KBNMiner is developed through the integration of cognitive maps and neural networks. To validate the effectiveness of the KBNMiner, an experiment with Web news information is conducted and its results are discussed.

Introduction

Recently, the number of large-scale repositories linked to the Web has been increased enormously. The expectations and dependencies on Web resources have also increased...
rapidly. Nevertheless, it is getting more difficult to collect and generate the useful information from the resources. Concerning this difficulty, the importance of data mining is being addressed more now than ever before, and many applications employing statistical and machine learning techniques are being developed (Changchien & Lu, 2001; Chiang, Chow, & Wang, 2000; Fayyad, Piatesky-Shapiro, & Smith, 1996a; Park, Piramuthu, & Shaw, 2001).

Although both techniques have their own strengths and weaknesses, machine learning has typically been employed in data mining applications (Bose & Mahapatra, 2001). Machine learning includes various techniques such as neural networks, rule induction, case-based reasoning, and genetic algorithms. These techniques are typically applied to analyze quantitative information rather than qualitative because the quantitative information is usually represented in the form of words or images. Qualitative information on the Web can be included primarily in a variety of text-based documents. As the most popular example, news on the Web can be considered; news typically consists of text and images. Unlike the quantitative information, in most cases, the qualitative information is not ready for use in data mining.

Among the techniques mentioned above, the neural networks have shown more considerable success in modeling financial data series (Refenes, Burgess, & Bentz, 1997; Walczak, 1999; Zhang et al., 1998). Neural networks have a flexible nonlinear function-mapping capability that can be used to approximate any continuous function with any desired accuracy (Cybenko, 1989; Hornik, Stinchcombe, & White, 1989). Neural networks are also nonparametric data-driven models that impose few prior assumptions on the underlying processes from which data are generated. This is an important advantage in financial data series forecasting. Neural networks have the ability to scan the data for patterns and can be used to construct nonlinear models. However, in financial data series forecasting, most neural network models are constructed on the basis of only quantitative information stored in database as the form of numeric, though qualitative information is able to improve the performance of the forecasting. For example, the qualitative information on the political situations, social conditions, international events, government policies, and traders’ psychology has influenced the movement of the financial data series (Kohara, Ishikawa, Fukuhara, & Nakamura, 1997; Kuo, Chen, & Hwang, 2001).

Considering this background, the following research questions may be addressed:

• How can the qualitative information be used in modeling financial data series?
• What is the effective tool in knowledge processing for utilizing the qualitative information in modeling financial data series?
• What is the effect of the qualitative information on the neural network performance when compared to the neural network model with only quantitative information?
• Is it more effective to train neural networks by separating the learning data into two groups on the basis of event information, which can be acquired from news information, than to apply traditional learning methods?
Related Content

Data Cleansing and Validation for Multiple Site Link Structure Analysis
[www.irma-international.org/chapter/data-cleansing-validation-multiple-site/31140/](www.irma-international.org/chapter/data-cleansing-validation-multiple-site/31140/)

Lightweight Collaborative Web Browsing
[www.irma-international.org/chapter/lightweight-collaborative-web-browsing/72951/](www.irma-international.org/chapter/lightweight-collaborative-web-browsing/72951/)

A Novel Approach to Construct Semantic Mashup using Patterns
[www.irma-international.org/article/a-novel-approach-to-construct-semantic-mashup-using-patterns/170369/](www.irma-international.org/article/a-novel-approach-to-construct-semantic-mashup-using-patterns/170369/)

Analyzing the Traffic Characteristics for Evaluating the Performance of Web Caching
[www.irma-international.org/chapter/analyzing-traffic-characteristics-evaluating-performance/53061/](www.irma-international.org/chapter/analyzing-traffic-characteristics-evaluating-performance/53061/)

Relay Selection in Distributed Transmission Based on the Golden Code Using ML and Sphere Decoding in Wireless Networks
[www.irma-international.org/chapter/relay-selection-distributed-transmission-based/72766/](www.irma-international.org/chapter/relay-selection-distributed-transmission-based/72766/)