Chapter 13
HONNs with Extreme Learning Machine to Handle Incomplete Datasets

Shuxiang Xu
University of Tasmania, Australia

ABSTRACT
An Extreme Learning Machine (ELM) randomly chooses hidden neurons and analytically determines the output weights (Huang, et al., 2005, 2006, 2008). With the ELM algorithm, only the connection weights between hidden layer and output layer are adjusted. The ELM algorithm tends to generalize better at a very fast learning speed: it can learn thousands of times faster than conventionally popular learning algorithms (Huang, et al., 2006). Artificial Neural Networks (ANNs) have been widely used as powerful information processing models and adopted in applications such as bankruptcy prediction, predicting costs, forecasting revenue, forecasting share prices and exchange rates, processing documents, and many more. Higher Order Neural Networks (HONNs) are ANNs in which the net input to a computational neuron is a weighted sum of products of its inputs. Real life data are not usually perfect. They contain wrong, incomplete, or vague data. Hence, it is usual to find missing data in many information sources used. Missing data is a common problem in statistical analysis (Little & Rubin, 1987). This chapter uses the Extreme Learning Machine (ELM) algorithm for HONN models and applies it in several significant business cases, which involve missing datasets. The experimental results demonstrate that HONN models with the ELM algorithm offer significant advantages over standard HONN models, such as faster training, as well as improved generalization abilities.

INTRODUCTION
Artificial Neural Networks (ANNs) have been providing significant benefits in many business applications. They have been actively used for applications such as bankruptcy prediction, predicting costs, forecast revenue, processing documents, and more (Kurbel, et al., 1998; Atiya, et al., 2001; Baesens, et al., 2003). Almost any neural network model would fit into at least one business area or financial analysis. Traditional statistical methods have been used for business applications with many limitations (Azema-Barac, et al., 1997; Blum, et al., 1991; Park, et al., 1993).
This chapter addresses using ANNs for handling business data for the following reasons. First, although usually considered a black-box approach, ANNs are a natural technology for data mining. ANNs are non-linear models that resemble biological neural networks in structure and learn through training. ANNs present a model based on the massive parallelism and the pattern recognition and prediction abilities of the human brain. ANNs learn from examples in a way similar to how the human brain learns. Then ANNs take complex and noisy data as input and make educated guesses based on what they have learned from the past, like what the human brain does. Given the requirements of data mining within large databases of historical data, ANNs are a natural technology for this application (McCue, et al., 2007). Next, ANNs (especially higher order ANNs) are able to handle incomplete or noisy data (Peng, et al., 2007; Wang, 2003). Databases usually contain noise in the form of inaccuracies and inconsistencies. Lack of data validation procedures may allow a user to enter incorrect data. Data can also become corrupt during migration from one system to another. Missing data is a common problem especially when data is collected from many different sources. Finally, ANNs hold superior predictive capability, compared with other data mining approaches (Xu, 2009; Zhang, et al., 2007; Fulcher, et al., 2006; Browne, et al., 2004; Kohonen, et al., 2000). The predictive accuracy of a data mining approach strongly influences its effectiveness and popularity. Higher predictive accuracy with real data is an obviously desirable feature.

Many of the important business data mining functions performed by ANNs are mirrored by those of the human brain. These include classification, clustering, associative memory, modeling, time-series forecasting, and constraint satisfaction (Cios, et al., 2007; Bigus, 1996). These tasks, which are important for human survival as a species, involve simultaneous processing of large amounts of data, where fast and accurate pattern recognition and responses are required. Classification refers to making distinctions between items, the most basic function performed by the human brain. We are able to analyse objects using the finest features to assess their similarities and differences. In the business environment, there is also a need for making classifications. Examples are: should a loan application for a new house be approved? Should an application for extending a line of credit to a growing business be approved? Should the new catalog of a company be mailed to this set of customers or to another set? All of these decisions are made based on classification. Clustering refers to the ability to group like things together. The business applications of clustering are mainly in the marketing arena. By clustering customers into groups based on similar attributes such as which products they buy or demographics they share, we can understand the markets in finer detail. Such information can be used to target specific groups of customers with products that many of them have previously purchased, or add-on services, which might appeal to the groups. Associative memory refers to associating two or more items. In business, many products are closely related to each other so when a customer purchases one of them he is likely to also buy the others. ANNs such as Bidirectional Associative Memories and Hopfield networks have been shown to be of such capabilities (Han, et al., 2006). Modeling refers to learning to predict outcomes based on existing examples. An experienced stock trader watches the changes of leading economic indicators to know when to buy or sell. With learning algorithms, ANNs are able to learn the existing examples and then, given new inputs, make predictions. Such ability to generalize on novel cases is one of the greatest strength of ANNs. An important variation of modeling is time-series forecasting, which looks at what has happened for some period back through time and predicts for some point in the future, a more difficult but more rewarding task. Finally, constraint satisfaction refers to solving complex problems that involve multiple simultaneous constraints. Having multiple con-