International Journal of Innovative Computing, Information and Control
Volume **1**, Number **4**, December **2005**

ICIC International ©2005 ISSN 1349-4198
pp. **635–642**

KNOWLEDGE DISCOVERY FOR LARGE DATA SETS USING ARTIFICIAL NEURAL NETWORK

Gangadhara T. Shobha, Sreenivasa C. Sharma

Department of Computer Science and Engineering
Rashtreeya Vidhyalaya College of Engineering
Mysore road, Bangalore 560-059, India
shobhatilak@rediffmail.com; rvrdscs@blr.vsnl.net.in

Doreswamy

Mangalore University
Mangala Gangothri, Mangalore 574-199, India
doreswamyh@yahoo.com

Received October 2004; revised May 2005

Abstract. *This paper deals with the application of artificial neural network for extracting knowledge of the predicted income for a prospective employee in US based on the already available database created from the population survey conducted in US. The survey reports the details of demographic and employment related characteristics of around 200,000 instances. A prediction module is developed using backpropagation multilayer perceptron (MLP) neural network to predict the income. A comparative study is undertaken for prediction of income for different learning rates. The validation results show that neural network achieved an acceptable level of performance in terms of maximum error being 0.000009045.*

Keywords: Backpropagation multilayer perceptron, Learning rate, Sigmoid activation, Neural network

1. **Introduction.** Artificial neural networks are a class of parametric, nonlinear statistical models that have found widespread use in many domains, including data mining, signal processing, medical diagnosis and finance. The typical network in such an application has 100-100,000 adjustable parameters and requires a similar number of training patterns in order to generalize well to unseen test data. Provided sufficient training data is available, the accuracy of the network is limited only by its representational power, which in turn is essentially proportional to the number of adjustable parameters. Thus, in domains where large volumes of data can be collected, such as stock market prediction, speech, face and character recognition, and web page classification - improved accuracy can often be obtained by training a much larger network [1,2,9,6]. Prediction is part of data mining the value of some attributes is predicted based on values already seen in the database [2].

In this paper a prediction module is developed to determine the income level of a person given his demographic characteristics and employment related details. The back propagation multi layered perceptron with various adjustable parameters is used to determine the income level accurately based on the input details provided.