ELSEVIER

Intelligent Data Analysis 2 (1998) 263–264

IDA
INTELLIGENT DATA ANALYSIS

www.elsevier.com/locate/ida

Letter From the Editor

Dear Colleague:

Welcome to Volume 2(4) of Intelligent Data Analysis journal!

When oil was discovered, many people called it "back gold". This has been proven to be correct since oil production has remained highly profitable. When terabytes of data that is generated in many enterprises are properly warhoused and used, we can treat it as a "pot of gold". Neither the real gold nor the black gold can be can be easily produced. Proper equipment and know-how are required for a successful discovery, operation and production. An analogy can be made for reaching the pot of gold. To reach it, we need to intelligently analyze our data aiming at the pot of gold. We can succeed the same way as the real gold and the black gold producers have succeeded for decades. The aim of our journal is to disseminate information that leads towards this discovery.

Volume 2(4) of IDA consists of five articles. The first article by Costa and Lerman explains the problem of binarizing categorical attributes that have a very large number of values. They introduce a method for extracting relevant predictive binary attributes from a large set of attributes. The extracted attributes could be used in any prediction methods that use binary attributes. A comparison between their method and CART is given in this paper. In the second article by Ray and Hsu, the authors investigate a from of modular neural network that is used for classification and is fed by externally distinct vector streams. They used the network for the classification of melodies presented as a direct audio events played by a human. Their results and observations showed that the technique was very robust and supported the theoretical deductions concerning the accuracy of the approach. Indurkhya and Weiss, in the third article of this issue, discuss the process of re-sampling techniques in decision trees and describe estimators of predictive performance for voted decision trees. These are classification trees that are generated from bootstrap or adaptive reasoning. The main advantage of these estimators, that are found by examining the performance of a single tree and its pruned sub-trees, is that they reveal the predictive potential of voted decision tress prior to applying computational procedures.

This issue of IDA also includes two application papers. The first one, by Camarinha-Matos and Martinelli, is on the application of machine learning techniques to a water distribution project. In this paper, the authors describe a machine learning subsystem that is embedded in a large system to control and manage water distribution networks. The goal of this subsystem is to extract knowledge from historical data and improve the performance of the entire water management system. Their preliminary results were quite promising. The second application paper by Humphrey, Cunningham and Witten is about knowledge visualization techniques for machine learning applications. In this paper, the authors analyze the questions that usually users of machine learning ask about models that are developed by

PII: S1088-467X(99)00009-8

learners. They demonstrate how users' questions can be mapped onto visualization tasks to create new graphical representations that show the flow of examples through a decision structure.

We always appreciate your feedback. We also welcome short conference overviews, workshop summaries and lessons learned from real world applications of data analysis techniques.

Best wishes,

A. Famili

Editor-in-Chief