Continuous Auditing and Data Mining

Edward J. Garrity
Canisius College, USA

Joseph B. O'Donnell
Canisius College, USA

G. Lawrence Sanders
State University of New York at Buffalo, USA

INTRODUCTION

Investor confidence in the financial markets has been rocked by recent corporate frauds and many in the investment community are searching for solutions. Meanwhile, due to changes in technology, organizations are increasingly able to produce financial reports on a real-time basis. Access to this timely information can help investors, shareholders, and other third parties, but only if this information is accurate and verifiable. Real-time financial reporting requires real-time or continuous auditing (CA) to ensure integrity of the reported information. Continuous auditing " is a type of auditing which produces audit results simultaneously, or a short period of time after, the occurrence of relevant events" (Kogan, Sudit, & Vasarhelyi, 2003, p. 1). CA is facilitated by eXtensible Business Reporting Language (XBRL), which enables seamless transmission of company financial information to auditor data warehouses. Data mining of these warehouses provides opportunities for the auditor to determine financial trends and identify erroneous transactions.

BACKGROUND

Auditing is a "systematic process of objectively obtaining and evaluating evidence of assertions about economic actions and events to ascertain the correspondence between those assertions and established criteria and communicating the results to interested parties" (Konrath, 2002, p. 5). In CA, the collection of evidence is constant, and evaluation of the evidence occurs promptly after collection (Kogan et al., 2003).

Computerized Assisted Auditing Techniques (CAATs) are computer programs or software applications that are used to improve audit efficiency. CAATs offer great promise to improve audits but have not met expectations "due to a lack of a common interface with IT systems" (Liang et al., 2001, p. 131). Also, "concurrent CAATS.... often require that special audit software modules be embedded at the EDP system design stage" (Liang et al., 2001, p. 131). Many entities are reluctant to allow the implementation of embedded audit modules, which perform CA, due to concerns these CAATs could adversely affect systems processing in areas such as reducing response times. Considering these difficulties, it is not surprising that continuous transaction monitoring tools are the second least used software by auditors (Daigle & Lampe, 2003).

These hurdles to CAAT usage are being minimized by the emergence of eXtensible Markup Language (XML) and eXtensible Business Reporting Language that minimize system interface issues. XML is a mark-up language that allows tagging of data to give the data meaning. XBRL is a variant of XML that is designed specifically for financial reporting and provides the capability of real-time online performance reporting. Both XML and XBRL enable the receiver of the data to seamlessly download information to the receiver's data warehouse

According to David & Steinbart (2000), data warehouses improve audit quality and efficiency by reducing the time needed to access data and perform data analysis. Improved audit quality should lead to early detection, and possible prevention, of fraudulent financial reporting. Auditor data warehouses may also be used in financial fraud litigations in providing evidence to evaluate the legitimacy of transactions and appropriateness of auditor actions in assessing transactions.

Data mining techniques are well suited to evaluate CA generated warehouse data but advances in audit tools are needed. Data mining and analysis software is the most commonly used audit software (Bierstaker, Burnaby, & Hass, 2003). Auditor data mining and analysis software typically includes low level statistical tools and auditor specific models like Benford's Law. Benford's Law holds that there is a naturally occurring pattern of values in the digits of a number (Nigrini, 2002). Significant variation from the expected number pattern may be due to erroneous or fraudulent transactions.

More robust auditing tools, using more sophisticated data mining methods, are needed for mining large databases and to help auditors meet auditing requirements. According to auditing standards (Statement on Audit Standards 99), auditors should incorporate unpredictability in procedures performed (Ramos, 2003). Otherwise, perpetrators of frauds may become familiar with common audit procedures and conceal fraud by placing it in areas that auditors are least likely to look.

MAIN THRUST

It is critical for the modern auditor to understand the nature of CA, and the capabilities of different data mining methods in designing an effective audit approach. Toward this end, this paper addresses these issues through discussion of CA, comparison of data mining methods, and we also provide a potential CA and data mining architecture. Although it is beyond the scope of this paper to provide an in-depth technical discussion of the details of the proposed architecture, we hope this stimulates technical research in this area and provides a starting point for CA system designers.

Continuous Auditing (CA)

Audits involve three major components: audit planning, conducting the audit, and reporting on audit findings (Konrath, 2002). The CA approach can be used for the audit planning and conducting the audit phases. According to Pushkin (2003), CA is useful for the strategic audit planning component that "addresses the strategic risk of reaching an inappropriate conclusion by not integrating essential activities into the audit plan" (p. 27). "Strategic information may be captured from the entity's Intranets and from the global internet using intelligent agents" (Pushkin, 2003, p. 28).

CA is also useful for performing the audit or what Pushkin (2003) refers to as the tactical component of the audit. "Tactical activities are most often directed at obtaining transactional evidence as a basis on which to assess the validity of assertions embodied in account balances" (p.27). For example, CA is useful in testing that entities comply with financial performance measures of debt covenants in loan agreements (Woodroof & Searcy, 2001).

CA requires prompt responses to high-risk transactions and the ability to identify financial trends from large volumes of data. Intelligent agents can be used to promptly identify and respond to erroneous transactions. Understanding the capabilities of data mining methods in identifying financial trends is useful in selecting an appropriate data mining approach for CA.

Comparing Methods of Data Mining

Data mining is a process by which one discovers previously unknown information from large sets of data. Data mining algorithms can be divided into three major groups: (1) mathematical-based methods, (2) logic-based methods, and (3) distance-based methods (Weiss & Indurkhya, 1998). Common examples from each major category are described below.

Mathematical-Based Methods

Neural Networks

An Artificial Neural Network (ANN) is a network of nodes modeled after a neuron or neural circuit. The neural network mimics the processing of the human brain. In a neural network, neurons are grouped into layers or slabs (Lam, 2004). An input layer consists of neurons that receive input from the external environment. The output layer communicates results of the ANN to the user or external environment. The ANN may also consist of a number of intermediate (hidden) layers. The processing of an ANN starts with inputs being received by the input layer and, upon being excited; the neurons are fired and produce outputs to the other layers of the system. The nodes or neurons are interconnected and they will send signals or "fire" only if the signals it receives exceed a certain threshold value. The value of a node is a non-linear, (usually logistic), function of the weighted sum of the values sent to it by nodes that are connected to it (Spangler, May, & Vargas, 1999).

Programming a neural network to process a set of inputs and produce the desired output is a matter of designing the interactions among the neurons. This process consists of the following: (1) arranging neurons in various layers, (2) deciding both the connections among neurons of different layers, as well as the neurons within a layer, (3) determining the way a neuron receives input and produces output (e.g., the type of function used), and (4) determining the strength of connections within the network by selecting and using a training data set so that the ANN can determine the appropriate values of connection weights (Lam, 2004).

Prior neural network research has addressed audit areas of risk assessment, errors and fraud, going concern audit opinion, financial distress, and bankruptcy prediction (Lin, Hwang, & Becker, 2003). Research has identified successful uses of ANN, however, there are still many other issues to address. For instance, ANN is effective for analytical review procedures although there is no clear guideline for the performance measures to use for this analysis (Koskivaara, 2004). Neural network research found differences between the patterns of quantitative