Decision Support Systems 84 (2016) 78–88

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

A knowledge based scheme for risk assessment in loan processing by banks

Partha Saha, Indranil Bose *, Ambuj Mahanti

Indian Institute of Management Calcutta, Diamond Harbour Road, Joka, Kolkata 700104, West Bengal, India

ARTICLE INFO

Article history:
Received 22 August 2015
Received in revised form 4 January 2016
Accepted 7 February 2016
Available online 13 February 2016

Keywords:
Knowledge-based systems
Compliance auditing
Deviation Pattern Components
Logistic regression
Risk management

ABSTRACT

Inadequacy in the compliance auditing (CA) process is one of the major reasons behind corporate frauds and accrual of non-performing assets within the banking sector. This phenomenon threatens the organization, stakeholders and society at large. The traditional CA process is slow and often inadequate in highly regulated and networked sectors such as banking, insurance and healthcare. This paper proposes a knowledge driven automated compliance auditing scheme for the processing of loans by banks. We collect 100 cases that are designated as fraudulent by banks and use them to design an automated risk score card model. The model uses text mining to automatically classify DPCs (Deviation Pattern Components) from unstructured text based cases. DPC patterns in a case give an early indication of the portfolio turning into a NPA. At the same time the cases are reviewed by five expert auditors in order to determine their risk level, risk impact and ease of detection. A logistic regression based model is used to derive risk scores of the case studies and classify the cases. By incorporating experts' opinion along with data mining techniques, the model automates the prediction of risk level, risk impact and ease of detection of fraudulent cases that deal with loan processing. The classifier performs well in terms of various performance metrics. The knowledge based method has the potential to save time and expensive human resources by automating the risk analysis of fraudulent loan processing cases reported by banks.

1. Introduction

The importance of risk based compliance auditing as one of the major components of corporate governance cannot be overemphasized as organizations throughout the world are increasingly facing a plethora of regulations for mandatory compliance. The severity of the enforcement of regulatory guidelines is the result of a series of scams and corporate frauds (e.g., Enron, Xerox, Parmalat, WorldCom, Lucent, Adelphia, Global Crossing, and Tyco) that have resulted in the loss of billions of dollars, an erosion of investor confidence and public mistrust in financial statements and the auditing processes [61]. This in turn has led to firms getting increasingly concerned about effective methods of fraud detection and prevention [3,41,57]. A survey conducted in 2014 has predicted that an organization loses 5% of its annual revenue on an average due to fraudulent incidents and the total loss due to fraud around the globe is over US$ 3.7 trillion [3]. These fraudulent incidents are particularly prevalent in the banking and other highly regulated and information intensive sectors (e.g., telecom, finance, insurance, healthcare) where fraud is seen as the most important cause for loss of revenue [29]. Fraud has become a complex phenomenon, and therefore it cannot be easily detected using known methods [4,5,21,22,44,66]. To this end researchers have identified, adopted and developed a plethora of systems, practices, methods and tools for detection of fraud, with and without the use of information technology [23,24,27,63]. IT enabled risk based auditing quantitatively measures misalignment between organizational work practices, statutory regulatory policies and guidelines. Stringent fines, including possible delisting of the company from the stock exchanges, potential jail terms for the errant management, investor mistrust and eventual bankruptcy of the organization are some of the consequences for failure to comply with auditing guidelines.

The role of the auditor in compliance auditing (a business process for checking the organization practices with the regulatory policies and guidelines of the business) is to identify and analyze any misalignment and non-compliance of the organization's rules and policies. The auditor's task includes verifying reliability of financial information and transaction, attesting security of the assets, loans and advances, and examining the financial statements and supporting records using sound auditing techniques. A distinct challenge in auditing is the repetitive, resource intensive nature of identifying non-compliant organizational issues. Manual verification of adherence to compliance rules is difficult and time consuming. Hence, the efficiency of the process is low and the cost is high due to the labor-intensive nature of compliance item generation, gathering of responses and subsequent analysis.

Despite the tremendous importance placed upon compliance auditing, due to its aforesaid manual nature, it faces severe limitations. Manual auditing is costly and time consuming. It is low in efficiency due to the labor intensive processes involving question answering, repeated checklist generation and analysis, and its repeatability is also low. The outcome of manual auditing is highly inconsistent and inaccurate due to subjectivity, human bias and error. Existing computer aided

* Corresponding author: Tel.: +91 33 2467 8300.
E-mail address: bose@iimcal.ac.in (I. Bose).

http://dx.doi.org/10.1016/j.dss.2016.02.002
0167-9236/

toolkits for compliance auditing mainly deal with lower levels of IT governance such as the management of configurations, patches and licenses. At the same time these products are vendor and technology specific, have poor reusability and are not capable of sharing knowledge.

To address this gap we aim to answer three research questions in the paper: (a) how can we construct a scheme for facilitating knowledge-based compliance auditing? (b) How do we measure its performance? (c) How does the proposed scheme improve upon standard manual auditing that is prevalent in the industry? The benefit accrued from the exercise is expected to improve the quality of auditing by reducing the costs and time involved in the process. We collect 100 case studies that deal with irregularities in loan processing in the SMEs (small and medium enterprises) sector. We use a text mining procedure to automatically identify anomalous patterns embedded in each case. Simultaneously, we seek the help of five reputed auditing experts in finding such patterns and listing the level of risk (L), the impact of risk (I) and the ease of detection of risk (D) for each of the cases. A single binary value (0 or 1) for L, I, and D is obtained on the basis of the expert auditors' evaluation using the principle of majority voting. These serve as output variables for a logistic regression model. The inputs to the model are the 0/1 values for each of the DPCs that indicate either the absence or presence of a specific DPC in a case study as detected by a text classifier. The logistic regression model learns through the training examples and predicts L, I, and D for the testing examples. When the performance of the model is ascertained by comparing the predicted values with the values obtained from the auditing experts, it shows a high degree of accuracy. The entire process automates the evaluation of a fraudulent case study in terms of L, I and D and determines a single risk score for the case study. When auditing experts are unavailable or too expensive for manual auditing or the time available for auditing of suspicious banking cases is limited, this logistic regression based model can mimic the performance of the experts with minimum time and cost.

The remainder of the paper is organized as follows: In Section 2 we examine literature related to risk based auditing. Section 3 introduces credit management in the small and medium enterprise sector of banking. The research method is discussed in Section 4. The results of the numerical experiments appear in Section 5. Discussion of the results and associated inferences is presented in Section 6. The limitations and the future scope of the work are described in Section 7. The paper concludes with Section 8.

2. Literature review

With the globalization of the economy, stakeholders' pressure on the financial sector to conform to ever increasing demands of regulations has considerably increased. Various security surveys emphasize the mandatory nature of compliance regulations in the light of various security scams around the world [9,20,28,58]. The abundance of country specific rules, regulations (Basel III, HIPPA, Gramm Leach Bliley), security standards, best practices (CobiT5, ISO 17799, ITIL, Baseline Protection Manual) and the mandatory cost of compliance to meet those regulations exert an enormous burden on organizations' manpower and scarce resources [8,56,67]. Inability to comply with various regulations attracts huge penalties from the regulators [56]. US based companies spend around $30 billion on corporate governance, risk management and compliance related solutions and this yearly expenditure continues to increase sharply [38,55]. Implementation of the labor intensive, continuous and iterative risk based auditing process requires alignment of organizational processes and internal control [70]. Hence, there is a need to develop dedicated information systems to significantly cope with and manage compliance with newer regulations [34,67,69]. Automation in the risk based auditing process is an ultimate goal for firms [2,68,70]. While full automation cannot be a reality due to current technological challenges and resource constraints, we can realistically aim for semi-automation in the auditing process. There is also a scope for valuable knowledge (explicit as well as implicit) sharing and reuse in auditing [65,69,71].

Analyzing fraudulent activities with technology enabled systems involves various types of decision support systems [42,54]. They provide insights into the complex multi-dimensional nature of frauds and render remedial actions or suggestions. The role of automated decision support systems in different countries and cultures have been examined by researchers [40]. In case of continuous auditing, changes in the concept of incident based auditing have been closely examined as well [5,6,7]. Embedding regulations by various stakeholders in the information systems architecture has been attempted in highly regulated industries [44]. Studies have also found that IT-enabled technologies are effective at brainstorming sessions for the detection of fraud [63].

Computer enabled e-auditing solutions to overcome the limitations of manual auditing processes have been in use for a few decades. Computer assisted auditing, on-time communication and reporting, and the use of ICT are some of the trends that require a computer assisted toolkit [60]. Accounts payable is also an area which requires computer assisted auditing due to fraud, fictitious vendor activity and a high volume of cash transactions [17]. One of the critical challenges, while migrating from legacy systems, has been the standardization of computing environments including data repository systems, data warehousing etc. [60]. Optimizing synchronized multidimensional database modules with integrated query processing and statistical modeling [51,60], implementing web enabled services for XML based accounting [50,51], and web based query processing [36] are some of the major challenges in this domain. Another challenge for real time auditing is the creation of a networked environment at the retail end (with an electronic funds collection and transfer system) that must be integrated with the inventory and accounting systems to enable online accounting [5,59].

Expert systems have been in use for computer assisted auditing since the last three decades [15,16]. Such systems include Automated Dynamic Audit Programme Tailoring (ADAPT) [35], KRisk of KPMG [11], WinProcess of Arthur Andersen, Planet of PwC and Audit Planning Advisor of Deloitte among others [15,16,74]. These systems have been primarily used to conduct repetitive tasks of low end procedural automation where decision making abilities are not important [26,39]. Excellent literature reviews of auditing expert systems are available for the interested readers and practitioners [1,26,39].

The success of expert systems has led to the designing of special purpose agents for auditing. A couple of special purpose auditing agents include EDGAR [52,53] (which enquires a user given firm's cash balance and computes financial ratios), FRAANK [14,43] (which acquires financial information and processes from the virtual web and parses and collates it for the purpose of auditing) and web based agents for Debt Covenant Compliance [72]. The use of agents has allowed automation in more complex transaction processing domains. The products described above address a small subset of the compliance management domain that includes change management, patch management, license management and configuration management. These solutions are also technology dependent as well as vendor specific [47,49,64].

XBRL (eXtensible Business Reporting Language) is used for the transfer of financial statements and is based on XML (eXtensible Markup Language) [10]. XBRL based systems have been used for multiple purposes like monitoring [51], standardization [60], analysis [14,43], information retrieval, linkages with client databases [13,46,72] and formulation of data standards in compliance with the GAAP taxonomy [75].

Neural networks have been widely used in auditing for a long time [30,31,32]. Its properties of learning and memory retention resemble that of the human brain and it closely follows human cognitive abilities. Its main strength lies in its high performance in prediction. A detailed summary of the use of neural networks in financial tasks [30,48], ratio analysis [19] and audit related research is available [17,45].

Our approach contains multiple techniques and ideas from diverse fields such as ontology, text mining, information retrieval, agent based systems, data mining and knowledge management. Ontology, a branch of metaphysics, is defined as "a formal specification of a shared conceptualization of a domain" [12]. Ontology captures implicit and explicit knowledge which can be shared, reused and consumed by autonomous agents [37]. At the University of Edinburgh, ontology and computational frameworks have been used in unison to support fraud detection, forensic analysis and legal reasoning [47]. The FF POIROT project in Belgium has attempted to build a detailed ontology of European law and financial fraud [12,33]. In this paper we develop a multilayered ontological architecture for e-auditing. While ontology is used at the top conceptual level to classify various layers in the audit scorecard, at the level of implementation we have used text mining and information retrieval [18] for segregating DPCs from unstructured text based cases, tabulated risk level, risk impact and risk detection and data mining techniques to tabulate risk scores to be shared across multiple layers. The differences between manual auditing, computer assisted auditing and knowledge driven auditing are shown in Table A1. A concise list of relevant literature in computer aided auditing is provided in Table A2. Both these tables appear in Appendix A.

3. The credit management process for SMEs

In this section we describe the underlying processes that support the construction of an audit risk scorecard for use by a bank. Even though the credit management process described here is quite generic, we focus on credit management for the SME sector.

3.1. Credit management in the SME sector in banks

A bank's credit management activity consists of six interlinked and integrated processes. They are: credit application, credit appraisal, credit sanction, credit disbursement, credit repayment and credit monitoring. The entire credit management process in the SME sector along with the activities of the concerned agents is shown in Appendix B.

3.2. Activities of banking agents in credit management for the SME sector

In the credit management operations in the SME sector there are seven principal agents. They are: customer (CU), bank staff (BS), bank management (BM), bank outsource agents (BOA), outside agent (OA, representing outside world), auditor (AU) and bank's Board of Directors (BBD). The last agent (bank's Board of Directors) is mainly concerned with policy issues of the bank and is excluded from the current discussion. While doing their activities, the various agents willingly or unwillingly commit many deviations from the prescribed rules and regulations laid down in the banking norms. The agents commit the deviations due to various complex circumstances including agents' group behavior, agents' pay off matrix, various incidents, internal state of the agents (including agents' mental state), among others. Table 1 summarizes the activities, sub-activities, compliance rules, violations and agents' involvement in loans processing for SMEs. The abbreviations used for describing the violation of compliance rules in Table 1 have been explained in Section 3.3.

3.3. Identifying DPCs of processes and agents in credit management for the SME sector

There are two types of violation patterns that can be observed in credit management in the banking industry. They are policy based violations and entity based violations. These DPCs help to identify vulnerable areas in the process of credit management. DPCs in a portfolio give an early indication of the portfolios turning into NPAs.

3.3.1. Policy based violations

The policy based violations consist of three DPCs that are described below:

Misuse of Power of Authority (*MOPA*): This type of fraud is normally committed by bank officials for personal/pecuniary benefits. Sanctioning a loan without undertaking the due diligence pre-sanctioning process, enhancing credit limit on fake mortgaged properties, sanctioning various loans to fake enterprises, fraudulent increase of overdraft amount, issue of duplicate cheque book without authorization, transferring money illegally to a dead/fictitious customer's account or a personal/relative's account and withdrawing later are some examples of MOPA.

Standard Operating Procedure Violations (*SOPVs*): These violations are committed by banking agents (internal or external) either due to the pressure of work or because of oversight in the process of meeting targets or by deliberate intent of violation. Sanctioning loans without proper documents, sanctioning loans without customer credentials (personal verification, address, income proofs), non-verification of credit worthiness of applicant (credit score) and non-performance of due diligence constitute SOPV.

Diversion of Loan (*DL*): A diversion of loan occurs when a loan obtained for one purpose is used for another purpose. Borrowers diverting the proceeds to other purposes or siphoning the loan to other accounts (other than the one for which it was originally sanctioned) constitute a diversion of a loan.

3.3.2. Entity based violations

The entity based violations consist of eight DPCs that are described below:

Over-Valuation of Collateral/Non-Existence of Collateral (*OVC/NEC*): This is a common deviation. Non-existence of collateral (movable or immovable property), non-existence of ownership (fictitious deed), multiple or partial ownership, litigated property, mortgaged collateral, collateral already sold out before mortgage etc. are examples that belong to this sub-category.

Fraudulent Documentation (*FD*): Fraudulent documents (title deeds, sales deeds, land ownership documents, land records, manipulated registration documents, stamp duties, revenue receipts, stamp paper, insurance paper, salary slips), genuine title deed of dead owner, invalid agreement of sale, fake know-your-customer (customer verification record), forged term deposit certificates, fabricated financial statements, fictitious purchase orders, fictitious export bills, spurious mutation certificates, fake physical inspection reports all constitute FD.

Fraudulent Instrumentation (*FI*): Forged cheques, double payment on cheques, photocopies or scanned copies of the original cheque, fake demand drafts, fake invoices, duplicate pass-books, fake withdrawal slips, forged signatures, spurious pay orders, fake letters of credit, duplicate insurance bonds etc. constitute FI.

Identity Theft (*IT*): Alteration of monetary instruments, fake signatures, false requests to alter specimen signatures, availing loans using fake employee certification, fake beneficiary account, obtaining instruments by using fake identities, loans obtained by fictitious firms etc. constitute IT.

Multiple Collusion (*MC*): Connivance between customer, bank officials and other parties (outside agents) to create fraudulent transactions (creating fake deeds, unauthorized sanctions, overvaluation, negating guidelines) to defraud the bank creates a scope for MC among different agents.

Table 1
Activities of agents in loans processing for SMEs.

Major activities of agents	Detailed sub-activities of agents	Compliance rules	Violation of compliance rules	Involvement of agents
Pre-sanction screening process for borrowers/guarantors	Verification (name, age, gender, identity, address and income proof)	Verifying authenticity of submitted documents	FD, SOPV	CU, BS, BM, BOA, OA, AU
Evaluation of borrower's credit scorecard	Loan amount, duration, down-payment, loan to property valuation ratio	Adherence to standard controls	FD, SOPV, MOPA	BS, BM, BOA, AU
Appraisal of borrower's credit history before loan sanctioning	Privileged information retrieval and sharing of borrower's credit history	Adherence to standard operating procedures	FD, SOPV	BS, BM, BOA, AU
Scrutiny of submitted information by loan applicants	Personal profile, bank account(s) details, financial information, sources of funds, repayment mode/period	Verifying authenticity of submitted documents	FD, SOPV	CU, BS, BM, BOA, OA, AU
Verification of scope and credentials of outside agents prior to empanelment	Past experience and competence, financial soundness and ability to service commitments, business reputation	Defining and executing criteria for outsourcing, performing due diligence	MOPA (external vendor induced fraud)	BM, BOA, AU
Review Service Level Agreement (SLA)	Formal SLA policy, SLA monitoring process, recourse for non-performance	Scrutinize SLA obligations for service contract breach	MOPA, SOPV	BM, BOA, AU
Incorporating Termination Clause (TC) for outsource agent	Termination clause/periods confidentiality/non-disclosure agreement	Scrutinize TC, obligations/liability	MOPA, SOPV	BM, BOA, AU
Verification of actual physical existence of collateral	Verification of actual physical existence of the properties/collateral	Opinion of empanelled lawyers, valuers etc.	OVC/NEC	CU, BS, BM, BOA, OA, AU
Multiple counter verification for high value loan portfolio	Verification of physical existence of the properties/collateral	Opinion of empanelled lawyers, valuers etc.	SOPV	CU, BS, BM, BOA, OA, AU
Counter checking of illegal transfer of mortgaged properties	Verification and ownership matching of borrower's properties/collateral	Opinion of empanelled lawyers, valuers etc.	FD, SOPV	CU, BS, BM, BOA, OA, AU
Undertake valuation for customer's mortgaged collateral	Verification of the properties/collateral being under/overvalued for mortgage	Opinion of empanelled lawyers, valuers etc.	FD, SOPV, OVC/NEC	CU, BS, BM, BOA, OA, AU
Verification of mortgage reused for procuring multiple loans	Verification and ownership matching of physical properties/collateral	Opinion of empanelled lawyers, valuers etc.	FD, SOPV, OVC/NEC	CU, BS, BM, BOA, OA, AU
Verify if the identical mortgage is resold without bank's sanction	Verification and ownership matching of physical properties/collateral	Opinion of empanelled lawyers, valuers etc.	FD, SOPV, OVC/NEC	CU, BS, BM, BOA, OA, AU
Verification of loan sanctioning authorities' jurisdiction limit	Verification of codified power of sanctioning authority	Codified power within banking rules and guidelines	MOPA, SOPV	BM
Verify if any agent exceeded its sanctioning power	Verification of codified power of sanctioning authority	Codified power within banking rules and guidelines	MOPA, SOPV	BS, BM, BOA, OA, AU
Penal provision against empanelled advocate/valuer/consultant for improprieties	Review SLA or invoke terminal clause penal provision if required	SLA, performance metric, terminal clause	MOPA, SOPV	BM, BOA, OA, AU
Monitoring loan accounts with irregular reimbursement and recovery	Daily branch level report, irregularities report, phone call, meetings, soft recovery measures, stressed asset recovery	Banking regulations for loan recovery, repayment, restructuring interest rate etc.	SOPV	CU, BS, BM, AU
Strategize dealings with NPA	Finding out degree (soft, medium, hard core) of NPA through customer interaction	Banking rules and regulations for loan recovery	SOPV	BM, BOA, OA, AU
Strategize for hard core NPA	Sending legal notice, filing money suit in courts, publishing names/photographs of defaulters, auctioning property, handing over difficult cases to recovery agencies	Banking rules and regulations for loan recovery and provision as well as legal and penal actions	SOPV	BM, BOA, OA, AU

FD (Fraudulent Documentation), FI (Fraudulent Instrumentation), SOPV (Standard Operating Procedure Violation), MOPA (Misuse of Power of Authority), OVC/NEC (Over-Valuation/Non-Existence of Collateral, DL (Diversion of Loan), IT (Identity Theft), MC (Multiple Collusion), AT (Account Takeover), TH (Theft), DF/CF (Digital/Cyber Fraud).

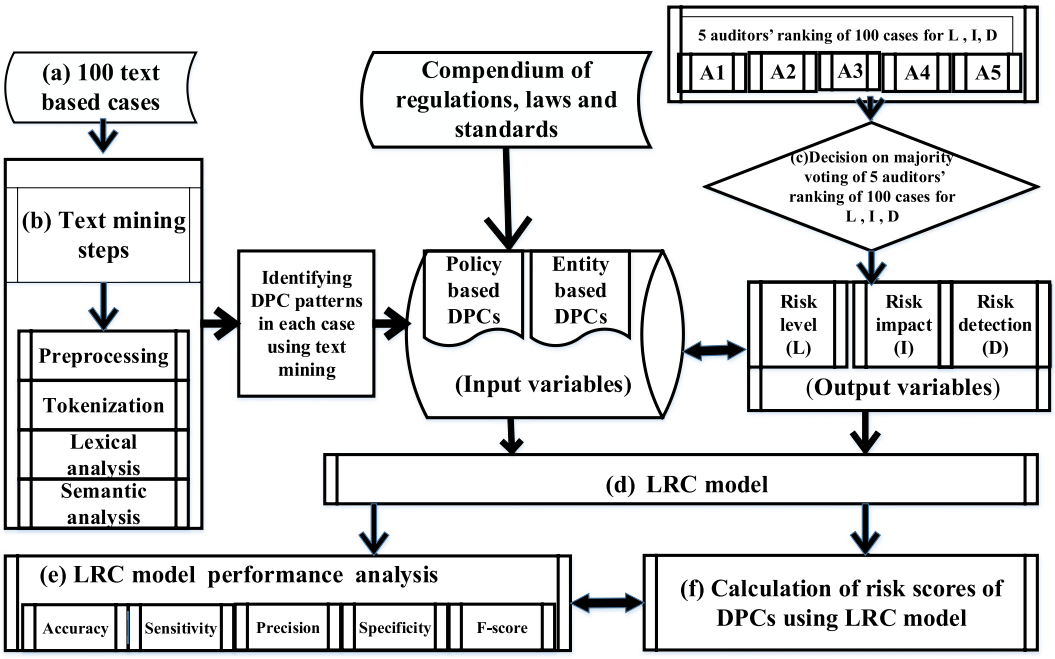

Fig. 1. A schematic diagram showing the different steps of the adopted research method.

Account Takeover (*AT*): Modifying books of accounts, transferring money from a customer account to a dead customer's account and withdrawing the balance, opening an account to appropriate funds meant for various government schemes, loan sanctions to fictitious people etc. constitute an account takeover.

Theft (*TH*): Stealing various monetary instruments constitutes theft. This normally occurs with dereliction of duty or connivance involving bank employees or bank management with fraudster(s) or outside agent(s). Normally these cases come to light after the perpetration of the actual crime.

Digital Fraud/Cyber Fraud (*DF/CF*): Misappropriating the user id and password of unsuspecting users and obtaining classified information through phishing attacks constitute digital fraud. In this case there is a strong involvement of inside and/or outside agents or fraudsters. In this particular case the gap in time between the cyber-attack and actual misappropriation of proceeds is very short.

4. Research method

In order to answer the three research questions that we listed in an earlier section, we resort to collection of data from the field and its analysis. Our goal is to answer the research questions through numerical evidence that can be gathered from a thorough analysis of field based data. We follow six steps in the analysis: (a) data collection, (b) identification of input variables, (c) identification of output variables, (d) task of classification, (e) evaluation of classification and (f) construction of risk score card. A schematic diagram showing the six steps of the adopted research method is shown in Fig. 1. The six steps are elaborated in details in the following paragraphs.

(a) Data collection: A portfolio of 100 cases that deal with fraudulent incidents in processing of loans by banks for SMEs is collected from various banks in an emerging economy. The authors collected case studies from ten public sector banks and three private banks of a developing country. Data have been systematically collected from various secondary sources (e.g., reports from banks, audit firms, newspaper reports, magazine articles, and public websites). Each bank maintains loan application records in their own format and no standard template exists across all banks. A number of middle and top ranking officials from two public sector banks were interviewed using a semi-structured questionnaire. Loan application reports from SMEs are considered in this research because the number and frequency of fraudulent activities in loan processing are very high for the SME sector. These cases are reports that are filed by the loan officer(s) of the bank after they suspect an anomaly in the loan application by the SME. The identity of the concerned banks and their customers is masked. Monetary instruments are converted from the local currency to USD (rounded to the nearest integer value). A descriptive summary of the cases is provided in Appendix C. A sample of five cases is provided in Appendix D.

(b) Identification of input variables: A text mining procedure is used to discover and extract patterns of DPCs (input variables) contained in each of the 100 cases. The text mining procedure splits the text into discrete words. Then it removes stop-words, articles, pronouns, interjections, prefixes and suffixes, normalizes spellings of words and detects sentence boundaries. In the next step, the procedure identifies the bag of words (BOW) for the identification and categorization of DPCs. We identify a set of keywords associated with each of the DPCs and put them in eleven buckets where each bucket represents a DPC. Initially the BOW per bucket is predetermined manually. When a new word is encountered in a case, the online thesaurus is used to find synonyms and then map the word to one of the buckets. After that the new word is added to the corresponding bucket. As a part of the procedure, a numeric weight is assigned to each word in the bucket to distinguish between important and unimportant words. Initially every word in the bucket is provided the same weight. Then a numeric weight is assigned to each word in the bucket according to the word's importance and degree of alignment within that particular bucket. The assignment of weight is useful in breaking ties when the words present in the case document are the same as words present in two different buckets. In that situation the case document will be assigned to that bucket where the relative importance of those words is higher. The cases contain different patterns of occurrences of DPCs. The text mining procedure identifies eleven triggers or DPCs from the various cases. The total number of DPCs is reduced using Principal Component Analysis. The two DPCs identity theft and theft shows strong correlation with

other independent variables and is removed from further consideration. After removing them, 9 DPCs are used for subsequent analysis.

(c) Identification of output variables: Five expert auditors are requested to examine each of the 100 cases and rate them in terms of risk level (L), risk impact (I) and ease of risk detection (D) (this is referred to subsequently as risk detection). In the Six Sigma model of calculating risk in lean manufacturing, risk is assumed to be made up of three components. Calculation of the Risk Priority Number (PPN) for performing a failure mode and effect analysis of a process or design is done by multiplying the three factors: risk occurrence (L), risk severity (I) and risk detection (D). RPN is used to sort processes or designs according to their riskiness [62,73]. Tables A9 to A14 appear in Appendix H. Table A9 provides a summary of the educational and professional accomplishments and activities of the five auditors and Table A10 lists the various thumb rules that are used by the auditors for determining their ratings. The correlation between the scores for the five auditors for evaluation of L, I, and D is shown in Tables A11, A12, and A13 respectively. The tables show that there is no strong correlation among the auditors' ratings. For each case, the values of L, I and D that are obtained from the auditors are discretized (high and low) from initial continuous values (on a scale of 10). This is done by selecting a cutoff value between 0 and 10. The average value of L for 100 cases computed by the five auditors is found to be 6.51. Similarly, the average values of I and D are 6.31 and 5.07 respectively. Hence, the cut-off values for L, I and D are chosen as 6.51, 6.31 and 5.07 respectively. For each case, if the value assigned by an auditor for L, I and D is found to be greater than or equal to the cut off value, then that value is converted to 1 (high), otherwise it is converted to 0 (low). A majority voting procedure is used to assign to each case a unique value (either 0 or 1). A summary of the auditors' ratings is included in Table A14. The table shows the frequency of occurrence of high value scores (scores beyond threshold values for L, I, and D) for different auditors. From the average frequency values that are reported in the table (near to 50) we may say that there is no false sense of high values for the three parameters and the judges' ratings are not overly skewed. These values are treated as output variables for the experiments to be run on the collected data. The data is organized into three tables. All three tables consist of 100 rows, each representing a fraudulent case. The columns contain input and output variables (values of all variables are 0 or 1). The input columns are made up of the nine DPCs (where 1 indicates the presence of a specific DPC in a case, 0 otherwise). Each table contains one column that represents the output. The output columns for the three tables indicate 'high' (1) or 'low' (0) values for L, I and D respectively.

(d) Task of classification: The LRC (logistic regression classifier) is selected to perform classification of L, I and D using data from the three tables respectively. LRC is a variation of the ordinary regression model when the predictors (i.e., the nine DPCs) as well as the response variables (i.e., L, I and D) are binary. The LRC is chosen over other classifiers as it satisfies two important pre-conditions: (i) it does not assume a linearity of relationship between the predictor and response variables, and (ii) it does not assume that response variables and error terms are distributed normally. These two conditions make the LRC a generic and powerful tool to deal with situations where the relationship between the predictors and response variables is not precisely known.

(e) Evaluation of classification: For each classification task, a 10-fold cross validation is used to eliminate bias and increase the robustness of the experiment. Several performance indicators like accuracy (ratio of correctly predicted records to total number of records), sensitivity or true positive rate of recall (proportion of positive cases which are correctly identified), specificity or true negative rate (proportion of negative cases which are correctly identified) and precision or measure of exactness are used to measure the performance of the classifier. The predictive performance is averaged over the 10-folds of cross validation.

(f) Construction of risk score card: The coefficients of DPCs are used to calculate the risk scores of individual DPCs. A risk score signifies the importance of an individual DPC in determining L, I and D in loan processing by the banks. The risk score of an individual DPC is calculated by multiplying three coefficients of that DPC (obtained for L, I and D respectively) with the frequency of occurrence of that particular DPC. In other words, risk score (S) of an individual DPC is the product of its frequency of occurrence (F), degree of severity and chance of detection (D). While risk severity measures the criticality of risk, risk detection measures the chance of detection of risky cases. Degree of severity is the product of risk level (L) and risk impact (I). Hence $S = F * L * I * D$. We also determined the 99% confidence interval for the risk score.

The risk scores of DPCs can be used as building blocks to calculate the risk weights of the different activities involved in processing of loans for the SMEs by banks. In loan processing there are several activities like pre-sanction of screening process for borrower's or guarantor's credentials, due diligence for the verification of credentials of outside agents prior to empanelment etc. For some portfolios, if these activities are not properly adhered to prior to loan disbursement, the loans may turn out to be NPAs. Hence, it is important to ascertain the risk of these activities quantitatively. Riskiness in an activity like pre-sanction of screening process for borrower or guarantor's credentials may be due to fraudulent documentation and/or violation of the standard operating procedure. Thus, the risk score of pre-sanction of screening process for borrower or guarantor's credential may be calculated by adding the risk scores of its two constituent DPCs i.e., fraudulent documentation and violation of standard operating procedure. Hence, the risk scores of DPCs obtained using the experimental procedure that is adopted by us can be used to construct the risk score card of various processes and sub-processes that are involved in disbursement of loans to SMEs.

5. Experimental results

In this section we examine the performance of the text classifier and the LRC in details. The text classifier is used to identify DPCs in each case while the LRC is used to enumerate the risk scores of DPCs. These risk scores are used to determine the order of importance of major activities performed by the agents for processing of loans. We determined the confidence interval of the risk score for each DPC. Thereafter several metrics (accuracy, sensitivity, specificity, and precision) are used for evaluating the performance of the LRC based model.

5.1. Performance of the text based classifier

In Table 2 we show the accuracy of the text mining procedure (how accurately the text mining procedure identifies DPCs from the 100 case studies). The performance of the text mining procedure is compared with that of the expert auditor. Appendix E shows the comparison between the text mining procedure and the expert auditors' identification of DPCs for five sample cases. An overall accuracy of 98.91% is achieved

Table 2
Analysis of performance of the text mining procedure.

Types of violation	DPCs	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Policy based violation	MOPA	98	89.47	100	100
	SOPV	100	100	100	100
	DL	96	94.03	100	100
Entity based violation	OVC/NEC	96	93.65	100	100
	FD	100	100	100	100
	FI	99	100	98.67	96.15
	IT	100	100	100	100
	MC	99	85.71	100	100
	AT	100	100	100	100
	TH	100	100	100	100
	DF/CF	100	100	100	100

The results represent the mean values obtained using the text mining procedure for 100 cases.

Table 3
Analysis of performance of the LRC.

Output variables	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Risk level	79(0.14)	93.81(0.08)	23.33(0.42)	83.17(0.15)
Risk impact	73(0.19)	90.83(0.11)	25(0.33)	77.43(0.18)
Risk detection	63(0.15)	35.79(0.27)	80.46(0.19)	55.90(0.39)

The results represent the mean and standard deviation (in brackets) obtained using a 10-fold cross-validation procedure.

for the text mining procedure. Similarly, high scores of sensitivity (96.62%), specificity (99.88%) and precision (99.65%) show the efficacy of the text mining procedure.

5.2. Performance of the LRC

Table 3 shows the performance of the LRC for classifying the three output variables L, I and D in terms of accuracy, error rate, sensitivity, specificity and precision. We observe that the LRC is able to classify risk level quite well with a reasonably high degree of accuracy. Since correctness in classification of high risk cases is more important than that for low risk cases, the high value obtained for sensitivity highlights the strength of the LRC for determining risk level. The same argument holds for the classification of risk impact as well. However, the LRC cannot classify risk detection so well and the accuracy is found to be lower than that for risk level and risk impact. Appendix F shows how the expert auditors' scores for L, I and D are converted to binary ratings for five sample cases.

5.3. Calculation of risk scores

From Table 4 we can observe that OVC/NEC (17.57) and DF/CF (6.7) are two DPCs that have a high risk. On the other hand, DL (−7.69) and FI (0.88) are DPCs that are found to be comparatively less risky. Table 4 also shows that there is substantial difference in the risk score between DPCs. We show the calculation of the risk score for five case studies in details in Appendix G.

5.4. Ranking of deviation meta-patterns based on risk scores of DPCs

We combined Tables 1 and 4 to construct Table 5 which combines major activities or processes undertaken by various agents (who are involved in loan processing by the bank for SMEs) with corresponding DPCs and determines the total risk scores associated with those activities. Subsequently we rank those nineteen activities in Table 5. We observe that three activities 'undertake valuation for customer's (borrower's) mortgaged properties or collateral', 'verification if the identical mortgage is not reused to obtain multiple loans from multiple sources' and 'verify if the identical mortgage is resold without sanction/knowledge/permission from the bank' have received the highest aggregate risk score of 28.01 whereas the activity 'due diligence for verification of credentials of outside agents prior to empanelment' has received the lowest aggregate risk score of 2.31.

6. Discussion

In this paper we examined three research questions: (a) how can we construct a scheme for facilitating knowledge-based compliance auditing? (b) How do we measure its performance? (c) How does the proposed scheme improve upon standard manual auditing that is prevalent in the industry? In the next three subsections we will examine the extent to which these objectives have been achieved in this research and the legitimate inferences which can be drawn from this research.

6.1. Constructing a framework for knowledge based auditing

Constructing a structured framework for knowledge based auditing is a challenging task. A human auditor uses his knowledge, learning, experience, judgment and understanding to arrive at a decision. However the same cannot be used in the knowledge based automated compliance auditing scheme. Here we have used text mining for identification of DPCs through keywords contained in real life cases. For each case we have tabulated the values of L, I and D (which are arrived at after taking a majority voting of the L, I, and D values obtained from five expert auditors). A LRC is used for measuring the performance of the classifier for prediction as well as for computing risk scores for DPCs. The risk scores of the DPCs are used to calculate the relative importance of the various processes in credit management for the SME sector. The entire chain of activities related to identification and quantification of risk scores is handled automatically through a scheme for knowledge based compliance auditing. When a bank receives an application for a loan from an SME and a manager flags it as suspicious, the knowledge based framework can be used to ascertain the risk level, risk impact and ease of detection of the fraudulent activity embedded in that application and a

Table 4
Risk scores of DPCs in loan management process.

DPC	Log of frequency of occurrence (lnF)	Log of risk level (lnL)	Log of risk impact (lnI)	Log of risk detection (lnD)	Risk score (S^)
MOPA	2.83	−0.83 (−2.93, 1.01)	−.03 (−2.07, 1.82)	0.33 (−1.25, 1.86)	2.31
SOPV	4.59	0.65 (−1.72, 3.08)	0.37 (−1.74, 2.42)	0.20 (−1.82, 1.99)	5.81
DL	4.22	0.29 (−5.04, 1.88)	0.52 (−4.39, 2.62)	−12.72 (−, −)	−7.69
OVC/NEC	4.16	1.07 (−1.49, 7.06)	−.21 (−2.49, 4.93)	12.55 (−, −)	17.57
FD	4.45	0.31 (−1.80, 2.40)	0.6 (−1.69, 2.74)	−0.73 (−2.67, 1.19)	4.63
FI	3.26	−2.03 (−4.60, 0.77)	−1.76 (−3.92, 0.72)	1.41 (−0.27, 2.97)	0.88
MC	1.95	2.40 (−2.75, 4.15)	2.14 (−3.03, 3.40)	−3.21 (−4.89, 1.88)	3.28
AT	3.76	1.40 (−1.83, 4.45)	0.95 (−1.62, 3.08)	−0.40 (−2.01, 1.54)	5.71
DF/CF	2.30	0.15 (−2.49, 2.67)	2.68 (−1.52, 4.12)	1.57 (−0.84, 3.45)	6.70

The risk score is obtained as: $S^{\wedge} = \ln S = \ln F + \ln L + \ln I + \ln D$.
Figures in brackets represent the 99% confidence interval for each value.

Table 5
Relative importance of activities of agents in loans processing for SMEs in terms of risk score.

Major activities or processes that are conducted by the agents	DPCs contained within those activities	Risk score of DPCs	Total risk score of activity	Ranking of activities in terms of risk
Pre-sanction screening process for borrowers/guarantors' credentials	FD	4.63	10.44	4
	SOPV	5.81		
Evaluation of borrower's credit scorecard	FD	4.63	12.75	3
	SOPV	5.81		
	MOPA	2.31		
Appraisal of borrower's credit history prior to loan sanctioning	FD	4.63	10.44	4
	SOPV	5.81		
Scrutinize/verify submitted information by the applicant for loan processing	FD	4.63	10.44	4
	SOPV	5.81		
Due diligence for verification of credentials of outside agents prior to empanelment	MOPA	2.31	2.31	7
Review Service Level Agreement (SLA) and performance metric	MOPA	2.31	8.12	5
	SOPV	5.81		
Incorporating Termination Clause (TC) while engaging with outsourcing agents	MOPA	2.31	8.12	5
	SOPV	5.81		
Physical verification of collateral	NEC	17.57	17.57	2
Counter verification of loan portfolio beyond a certain value	SOPV	5.81	5.81	6
Verification/counter checking if mortgaged properties are transferred illegally	FD	4.63	10.44	4
	SOPV	5.81		
Undertake valuation for customer's (borrower's) mortgaged properties or collateral	FD	4.63	28.01	1
	SOPV	5.81		
	NEC	17.57		
Verification if the identical mortgage is not reused to obtain multiple loans from multiple sources	FD	4.63	28.01	1
	SOPV	5.81		
	NEC	17.57		
Verify if the identical mortgage is resold without sanction/knowledge/permission from the bank	FD	4.63	28.01	1
	SOPV	5.81		
	NEC	17.57		
Verify if loan sanctioning process/amount is within sanctioning authorities' jurisdiction/ permissible limit	MOPA	2.31	8.12	5
	SOPV	5.81		
Verify if any agent exceeded its sanctioning power	MOPA	2.31	8.12	5
	SOPV	5.81		
Undertake penal provision against bank's empanelled advocate/valuer/consultant for professional improprieties or dereliction of duties	MOPA	2.31	8.12	5
	SOPV	5.81		
Provision of close monitoring of loan accounts with irregular reimbursement and recovery	SOPV	5.81	5.81	6
Strategize dealings with NPA	SOPV	5.81	5.81	6
Strategize for hard core NPA	SOPV	5.81	5.81	6

composite risk score can be calculated on the basis of that. If the risk score is found to be sufficiently high the case may be referred to a deeper investigation. Since a bank receives a large number of loan applications from SMEs and many of them can be flagged as suspicious, this automated method can help in directing only the most severe cases toward being investigated, thereby allowing better usage of the expert human resources available to the bank and leading to a related savings in cost.

6.2. Measuring the performance of a knowledge driven auditing system

Measuring the performance of a knowledge driven compliance auditing system is a complex task. We measure the performance of the compliance auditing system in terms of the performance of its sub-processes i.e. predictive performance of the text mining procedure and the LRC, assessment of risk scores of DPCs and subsequent risk scores of various processes involved in the credit management of SMEs.

Table 2 lists the accuracy, sensitivity, specificity, precision etc. of the text mining procedure which is used on 100 cases to accurately predict DPC patterns. The text mining procedure demonstrates a high level of accuracy (98.91%). While identifying DPCs that are contained in a case, the text mining procedure provides an explicit representation of the knowledge entities. The scope of decoupling of domain knowledge from model architecture does not exist in manual auditing but the text mining procedure is able to perform the task without any intervention from the human auditor. Our result compares favorably with the highest accuracy (82–83%) obtained using data mining models in related domains such as prediction of default of customers and prediction of financial distress among firms. A list of benchmark papers and the accuracies obtained therein are reported in Appendix I. Table 3 illustrates the efficiency of the LRC. The LRC exhibits an accuracy of 79% for classification of risk level as seen from Table 3, although accuracies in classification of risk impact (73%) and risk detection (63%) are relatively poor. Binary values of L, I, and D are obtained from domain experts and a majority voting procedure is used to compute the final values of L, I, and D for each case. The LRC model is built by taking DPCs as input variables and L, I, and D values that are obtained from the expert auditors as output variables. It is built on the assumption that the voting procedure leads to elicitation of correct values of L, I and D from the experts.

In Table 4 we calculate the risk scores of each of the nine DPCs. From this table we observe that OVC/NEC (17.57) and DF/CF (6.70) are the two most vulnerable areas in loan processing. Our finding is similar to that of a recent survey where OVC/NEC is reported to be the second most prevalent form of fraud in retail banking [25]. This calls for a modification of banking business processes and optimal resource reallocation including upgrading of cyber security to prevent fraudulent activities. Similarly, DL (−7.69) and FI (0.88) rank very low in terms of vulnerability. It implies that fraudulent activities are more prevalent in the lower levels of loan processing activities than in the higher echelons.

Table 5 tabulates fraudulent activities of the various processes in loan management in the SME sector. Here too, fraudulent activities that are related to collateral, borrower's credit scores, submitted documents are found to be highly vulnerable. We identify the most potent areas of vulnerabilities and detect patterns of anomalies in the organization's process flow. This knowledge is useful for a loan granting bank. Manual auditing often suffers from duplication of effort that leads to misutilization of scarce resources and results in low efficiency.

Sharing and reuse of procedures and findings can minimize this wastage. The methods that are used and the results that are obtained in Tables 2 and 3 can be reused as they are generic.

6.3. Comparative performance analysis of knowledge driven auditing with manual auditing

Knowledge based auditing may be compared with human auditing as well as computer aided auditing along multiple dimensions. In terms of cost, repeatability and time consumption, a knowledge based model performs very well in comparison to human auditing. Knowledge based auditing is also amenable to automation and is free from subjectivity and human bias. While the human auditor makes inferences on the basis of only a small fraction of sample data, knowledge based audit methods can use the full data set to make suitable inferences.

Multiple computer aided technologies have been used for auditing including expert systems, XML/XBRL, artificial computing agents, belief functions, real time based accounting, data based system etc. As stated previously, the currently available auditing expert systems mostly address the automation of low level repetitive tasks without complex decision making capabilities [1,15,26]. Various expert systems that are used in accounting firms represent these trends through the automation of lower level audit processes [11,16,35,74]. This contrasts sharply with the approach that is adopted in this paper which automates the discovery of DPCs that are contained in the case studies through a text mining procedure and determines risk scores for DPCs. These tasks involve complex decision making. Unlike computer-aided auditing, which requires prior standardization of data sources and can only be applied to processes that are fully automated [59,60], our research method assumes no such restriction and the text mining procedure can be applied on unstructured and descriptive case studies. Unlike human auditing both computer-aided and knowledge based auditing work on a real time basis [4,5], thus making them look like an auditing motion picture in comparison to an accounting still photograph that is provided by the manual process of auditing.

The proposed system described in the paper can be easily integrated to the backend of an application server. In order to do that, the bank will need to design a standard online template that will allow the receiving officer to fill in numerical details as well as textual comments about the application for loan. The report generated from the use of this template can serve as the input to the proposed system. Based on the description received, the system will compute the risk score for the application. The loan official can then either reject a problematic application whose risk score is very high or scrutinize it further and make a request for further information. This will help to save critical manpower and minimize the scope of human error.

The benefits that are accrued from knowledge based auditing involve improving preventive controls and taking corrective actions (including business process reengineering) as per the advice of the auditing system. It also impacts the decision making process related to granting of loans for SMEs. The resulting benefits that accrue include increase in efficiency, consistency, accuracy, reusability and repeatability of the risk based auditing process. Timely detection of risky portfolios through automated models minimizes operational risks.

7. Limitations and future scope of work

One limitation of the paper is that the findings that are reported in this research can be a result of the specific case studies that are included in our portfolio of case studies. A full evaluation of the benefits of our knowledge driven audit score card model will require comparative evaluation of many more case studies. While we have resorted to analysis of 100 case studies from an emerging economy, future research can look into evaluation of such case studies from multiple geographical locations in order to obtain better generalizability of results. While we have restricted ourselves to the study of processing of loans for SMEs by a bank in future it may be possible to use the audit scorecard model for constructing ontologies that are sharable across multiple domains (e.g., risk based auditing for telecommunication, healthcare, airlines and other services sectors). It remains to be seen to what extent there will be similarities as well as differences in the construction of ontologies for use in different domains.

Another limitation of the research method adopted by us is that it is totally dependent on the quality of the experimental data (i.e., textual records of the activity that is recorded). If the textual description is faulty our analysis will lead to erroneous results. This can be improved if banks use a standard template for reporting such cases. Another future challenge to be taken up is related to missing, unknown, ambiguous and misleading data (especially related to fraudulent cases) that are usually present in case studies. Building the capability for fault tolerance is not within the scope of this research. However it may be possible to handle this issue through the use of fuzzy logic and weighted fuzzy production rules for incorporating approximation in reasoning (while handling missing/ambiguous/imprecise knowledge relationships) during the construction of a semantically enabled risk based auditing tool.

We have used the LRC for predicting the risk level, risk impact and risk detection. Future researchers may want to use some other classifiers like Classification Tree, Naive Bayes, Neural Network and Support Vector Machine for performing the same task. While the LRC is a logical choice in this paper, it remains to be seen whether choosing other classifiers or using an ensemble of classifiers can improve the performance of the classification process.

8. Conclusion

Banks process numerous applications for loans from SMEs. Many of these applications are often fraudulent. In order to appropriately assess these applicants banks resort to expert auditors who evaluate the riskiness of these loan applications. This is a time consuming and expensive process. Our research attempts to solve this practical problem through a knowledge based scheme. Based on the textual description of the case under consideration our method builds a knowledge base that relates the occurrence of DPCs in a case study to its risk level, risk impact and ease of detection. We use text mining to discover important keywords that identify the DPCs and at the same time we use five expert auditors to rate the cases. Using a logistic regression based classification model we are able to relate the DPCs to the risk assessment of the cases. The model built by us is able to predict the risk characteristics of a case study based on the presence of specific DPCs with a high degree of accuracy.

This research contributes to the literature on knowledge based auditing. It provides a scheme that encapsulates the domain knowledge of expert auditors by linking them to occurrences of DPCs in fraudulent applications. The model that is built by us is reusable and can be used to obtain the risk score of a spurious case automatically without the help of an expert auditor. The method proposed in this research is easy to implement and is likely to save a bank time and money in processing loan applications. In order to fine tune the model and make it more accurate and robust, extensive experiments using data sets from different types of banks located in disparate geographies may be conducted in the future.

Acknowledgment

The third author gratefully acknowledges financial support received from the Indian Institute of Management Calcutta for the research project “Towards an Automated Adaptive Compliance Auditing System” (R1109, 019/RP:TAACAS/3483/11-12).

Appendix A. Supplementary material

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.dss.2016.02.002.

References

[1] M.J. Abdolmohammadi, Decision support and expert systems in auditing: a review and research directions, Accounting and Business Research. 17 (66) (1987) 173–185.
[2] S.N.H. Abdullah, M. Induslka, S. Shazia, A study of compliance management in information systems research, Proceedings of the European Conference on Information Systems, Verona, Italy, 2009 (http://aisel.aisnet.org/ecis2009/5).
[3] ACFE, Report to the Nations on Occupational Fraud and Abuse, Association of Certified Fraud Examiners, Austin, Texas, USA, 2014 (http://www.acfe.com/rttn/docs/2014-report-to-nations.pdf).
[4] M. Alles, A. Kogan, M.A. Vasarhelyi, Feasibility and economics of continuous assurance, Auditing: A Journal of Practice and Theory 21 (1) (2002) 125–138.
[5] M. Alles, A. Kogan, M.A. Vasarhelyi, Restoring auditor credibility: tertiary monitoring and logging of continuous assurance systems, International Journal of Accounting Information Systems 5 (2) (2004) 183–202.
[6] M. Alles, G. Brennan, A. Kogan, M.A. Vasarhelyi, Continuous monitoring of business process controls: a pilot implementation of a continuous auditing system at Siemens International, Journal of Accounting Information Systems 7 (2) (2006) 137–161.
[7] M.G. Alles, A. Kogan, M.A. Vasarhelyi, Putting continuous auditing theory into practice: lessons from two pilot implementations, Journal of Information Systems 22 (2) (2008) 195–214.
[8] H. Ashbaugh-Skaife, D. Collins, W.R. Kinney Jr., R. LaFond, The effect of SOX internal control deficiencies and their remediation on accrual quality, The Accounting Review 83 (1) (2008) 217–250.
[9] AusCERT, Australian Computer Crime and Security Survey, 2006 (http://www.auscert.org.au/images/ACCSS2006.pdf).
[10] A.A. Baldwin, C.E. Brown, B.S. Trinkle, XBRL: an impacts framework and research challenge, Journal of Emerging Technologies in Accounting 3 (1) (2006) 97–116.
[11] T.B. Bell, C.B. Jean, M.J. Karla, F.S. Edward, KRisk: a computerized decision aid for client acceptance and continuance risk assessments, Auditing: A Journal of Practice & Theory 21 (2) (2002) 97–113.
[12] W.N. Borst, Construction of Engineering Ontologies for Knowledge Sharing and ReusePhD Thesis The University of Twente, Netherlands, 1997 ISSN: 1381-3617 (CTIT Ph. D-series No. 97-14). (http://doc.utwente.nl/17864/1/t0000004.pdf).
[13] A.F. Borthick, D.R. Jones, R. Kim, Developing database query proficiency: assuring compliance for responses to web site referrals, Journal of Information Systems 15 (1) (2001) 35–56.
[14] M. Bovee, A. Kogan, K. Nelson, R.P. Srivastava, M.A. Vasarhelyi, Financial reporting and auditing agent with net knowledge (FRAANK) and extensible business reporting language (XBRL), Journal of Information Systems 19 (1) (2005) 19–41.
[15] C.E. Brown, D.S. Murphy, The use of auditing expert systems in public accounting, Journal of Information Systems 4 (2) (1990) 63–72.
[16] C.E. Brown, Expert systems in public accounting: current practice and future directions, Expert Systems with Applications 3 (1) (1991) 3–18.
[17] T.G. Calderon, J.J. Cheh, A roadmap for future neural networks research in auditing and risk assessment, International Journal of Accounting Information Systems 3 (4) (2002) 203–236.
[18] M. Christopher, R. Prabakhar, S. Heinrich, Introduction to Information Retrieval, Cambridge University Press, NY, USA, 2008.
[19] J.R. Coakley, C.E. Brown, Artificial neural networks applied to ratio analysis in the analytical review process, International Journal of Intelligent Systems in Accounting, Finance and Management 2 (1) (1993) 19–39.
[20] C. Cockrell, D.N. Stone, Team discourse explains media richness and anonymity effects in audit fraud cue brainstorming, International Journal of Accounting Information Systems 12 (3) (2011) 225–242.
[21] C.P. Cullinan, S.G. Sutton, Defrauding the public interest: a critical examination of reengineered audit processes and the likelihood of detecting fraud, Critical Perspectives on Accounting 13 (3) (2002) 297–310.
[22] J.S. Davis, H.L. Pesch, Fraud dynamics and controls in organizations, Accounting, Organizations and Society 38 (6–7) (2007) 469–483.
[23] R. Debreceny, S.-L. Lee, W. Neo, J.S. Toh, Employing generalized audit software in the financial services sector: challenges and opportunities, Managerial Auditing Journal 20 (6) (2005) 605–618.
[24] R.S. Debreceny, G.L. Gray, Data mining journal entries for fraud detection: an exploratory study, International Journal of Accounting Information Systems 11 (3) (2010) 157–181.
[25] Deloitte, Indian Banking Fraud Survey: Navigating the Challenging Environment, 2012 (http://www.deloitte.com/assets/DcomIndia/Local%20Assets/Documents/Thoughtware/India_Banking_Fraud_Survey_2012.pdf).
[26] E.L. Denna, J.V. Hansen, R.D. Meservy, Development and application of expert systems in audit services, IEEE Transactions on Knowledge and Data Engineering 3 (2) (1991) 172–184.
[27] M.M. Eining, D.R. Jones, J.K. Loebbecke, Reliance on decision aids: an examination of auditors' assessment of management fraud, Auditing: A Journal of Practice & Theory 16 (2) (1997) (Fall 1997).
[28] Ernst & Young, Global Information Security Survey (2013), 2013 (http://www.ey.com/Publication/vwLUAssets/EY_2013_Global_Information_Security_Survey/$FILE/EY-GISS-Under-cyber-attack.pdf).
[29] P.A. Estévez, C.M. Held, C.A. Perez, Subscription fraud prevention in telecommunications using fuzzy rules and neural networks, Expert Systems with Applications 31 (2) (2006) 337–344.
[30] A. Fadlalla, C.H. Lin, An analysis of the applications of neural networks in finance, Interfaces 31 (4) (2001) 112–122.
[31] K. Fanning, K.O. Cogger, R. Srivastava, Detection of management fraud: a neural network approach, International Journal of Intelligent Systems in Accounting, Finance & Management 4 (2) (1995) 113–126.
[32] K. Fanning, K.O. Cogger, Neural network detection of management fraud using published financial data, International Journal of Intelligent Systems in Accounting, Finance & Management 7 (1) (1998) 21–41.
[33] Semantics Technology and Applications Research Laboratory (STAR Lab). FF POIROT. (http://www.ffpoirot.com/default.htm).
[34] J. Fisher, Compliance in the performance management context: what technologies could simplify compliance and automate information gathering? Bank, Accounting & Finance 20 (4) (2007) 41–49.
[35] P. Gillett, Automated dynamic audit programme tailoring: an expert system approach, Auditing: A Journal of Practice and Theory 12 (1) (1993) 173–189.
[36] S.M. Groomer, U.S. Murthy, Continuous auditing of database applications: an embedded audit module approach, Journal of Information Systems 3 (2) (1989) 53–69.
[37] M. Gruninger, J. Lee, Ontology applications and design, Communication of the ACM 45 (2) (2002) 39–41.
[38] J. Hagerty, B. Kraus, GRC in 2010: $29.8B in Spending Sparked by Risk, Visibility, and Efficiency. Boston, MA, USA, 2009 (http://www.cmo-hseq-software.com/about_cmoglobal/AMR-GRC-in-2010.pdf).
[39] J.V. Hansen, W.F. Messier, Expert systems in auditing: the state of the art, Auditing: A Journal of Practice & Theory 7 (1) (1987) 94–105.
[40] E. Huerta, T. Glandon, Y. Petrides, Framing, decision-aid systems, and culture: exploring influences on fraud investigations, International Journal of Accounting Information Systems 13 (4) (2012) 316–333.
[41] M. Jans, N. Lybaert, K. Vanhoof, Internal fraud risk reduction: results of a data mining case study, International Journal of Accounting Information Systems 11 (1) (2010) 17–41.
[42] A. Kogan, E.F. Sudit, M.A. Vasarhelyi, Continuous online auditing: a program of research, Journal of Information Systems 13 (2) (1999) 87–103.
[43] A. Kogan, K.M. Nelson, R.P. Srivastava, M.A. Vasarhelyi, M. Bovee, Design and Applications of an Intelligent Financial Reporting and Auditing Agent With Net Knowledge (FRAANK), The University of Kansas, 2002 1–37 (https://kuscholarworks.ku.edu/dspace/handle/1808/141).
[44] V. Kartseva, J. Hulstijn, J. Gordijn, Y.-H. Tan, Control patterns in a health-care network, European Journal of Information Systems 19 (3) (2010) 320–343.
[45] E. Koskivaara, Artificial neural networks in analytical review procedures, Managerial Auditing Journal 19 (2) (2004) 191–223.
[46] G. Lau, K.H. Law, G. Wiederhold, Legal information retrieval and application to e-rule making, Proceedings of the Tenth International Conference on Artificial Intelligence and Law, Bologna, Italy 2005, pp. 146–154 (http://pdf.aminer.org/000/284/091/what_you_saw_is_what_you_want_using_cases_to.pdf).
[47] R. Leary, W. Vandenberghe, J. Zeleznikow, Towards a financial fraud ontology: a legal modelling approach, Proceedings of the Tenth International Conference on Artificial Intelligence and Law, Edinburgh, UK, 2003.
[48] J.W. Lin, M.I. Hwang, J.D. Becker, A fuzzy neural network for assessing the risk of fraudulent financial reporting, Managerial Auditing Journal 18 (8) (2003) 657–665.
[49] Managesoft, Managesoft Compliance Manager, 2007 (http://www.managesoft.com/product/compliance/index.xml).
[50] U.S. Murthy, An analysis of the effects of continuous monitoring controls on e-commerce system performance, Journal of Information Systems 18 (2) (2004) 29–47.
[51] U.S. Murthy, S.M. Groomer, A continuous auditing web services model for XML-based accounting systems, International Journal of Accounting Information Systems 5 (2) (2004) 139–163.
[52] K.M. Nelson, A. Kogan, P.P. Srivastava, M.A. Vasarhelyi, Virtual auditing agents: the EDGAR agent example, Proceedings of the Hawaii International Conference on System Sciences 1998, pp. 396–404 (Hawaii, USA).
[53] K.M. Nelson, A. Kogan, R.P. Srivastava, M.A. Vasarhelyi, H. Lu, Virtual auditing agents: the EDGAR agent challenge, Decision Support Systems 28 (3) (2000) 241–254.
[54] L. Nelson, Stepping into continuous audit, Internal Auditor 61 (2) (2004) 27–29.
[55] OpenPages, Risk Management Investments to Rise in 2010, 2009 (http://www.marketwired.com/press-release/risk-management-investments-to-rise-in-2010-openpages-survey-reveals-1204523.htm).
[56] E. Parry, SOX Wars: CIOs Share Ideas, Fears on Sarbanes–Oxley Compliance, 2004 (http://searchcio.techtarget.com/news/994763/SOX-Wars-CIOs-share-ideas-fears-on-Sarbanes–Oxley-compliance).
[57] PricewaterhouseCoopers, Cybercrime: Protecting Against the Growing Threat — Global Economic Crime Survey, London, UK, 2011 (http://www.pwc.com.br/pt/publicacoes/assets/pesquisa-crimes-digitais-11-ingles.pdf).
[58] PricewaterhouseCoopers, Information Security Breaches Survey 2014, 2014 (http://www.pwc.co.uk/assets/pdf/cyber-security-2014-technical-report.pdf).
[59] Z. Rezaee, W. Ford, R. Elam, Real-time accounting systems, Internal Auditor 57 (2) (2000) 62–67.
[60] Z. Rezaee, A. Sharbatoghlie, R. Elam, P.L. McMickle, Continuous auditing: building automated auditing capability, Auditing: A Journal of Practice and Theory 21 (1) (2002) 147–163.
[61] Z. Rezaee, Restoring public trust in the accounting profession by developing anti-fraud education, programs, and auditing. Managerial, Auditing Journal 19 (1) (2004) 134–148.

[62] S.J. Rhee, K. Ishii, Using cost based FMEA to enhance reliability and serviceability, Advanced Engineering Informatics 17 (2003) 179–188.
[63] A.L. Smith, U.S. Murthy, T.J. Engle, Why computer-mediated communication improves the effectiveness of fraud brainstorming, International Journal of Accounting Information Systems 13 (4) (2012) 334–356.
[64] Symantec, Improving IT Compliance: Guidance for Midsize Organizations, 2006 (http://i.i.cbsi.com/cnwk.1d/html/itp/Symantec_Improving_IT_Midsize.pdf).
[65] R.A. Teubner, T. Feller, Information technology, governance und compliance, Wirtschaftsinformatik 50 (5) (2008) 400–407.
[66] C.S. Throckmorton, W.J. Mayew, M. Venkatachalam, L.M. Collins, Financial fraud detection using vocal, linguistic and financial cues, Decision Support Systems 74 (2015) 78–87.
[67] L. Volonino, G.H. Gessner, G.F. Kermis, Holistic compliance with Sarbanes–Oxley, Communications of the Association for Information Systems 14 (11) (2004) 219–233.
[68] J.A. Wheeler, Hype Cycle for Governance, Risk and Compliance Technologies, 2014 (https://www.gartner.com/doc/2801317/hype-cycle-governance-risk-compliance).
[69] M. Wiesche, M. Schermann, H. Krcmar, Exploring the contribution of information technology to governance, risk, and compliance (GRC) initiatives, Proceedings of the Nineteenth European Conference on Information Systems, Helsinki, Finland, 2011.
[70] A.K.Y. Wong, F. Yip, P. Ray, N. Paramesh, Semantic data integration for IT governance, Proceedings of the First International Workshop on Semantic e-Science, PRC, Beijing, 2006.
[71] A.K.Y. Wong, F. Yip, P. Ray, N. Paramesh, Towards semantic interoperability for IT governance: an ontological approach, Computing and Informatics Journal 27 (2008) 1001–1025.
[72] J. Woodroof, D. Searcy, Audit Implications of Internet Technology: Triggering Agents Over the Web in the Domain of Debt Covenant Compliance. In Proceedings of the Thirty-fourth Hawaii International Conference on System Sciences, Hawaii, USA, 2001.
[73] N. Xiao, H.Z. Huang, Y.L. Liping, Multiple failure modes analysis and weighted risk priority number evaluation in FMEA, Engineering Failure Analysis 18 (2011) 1162–1170.
[74] N. Zhao, D.C. Yen, I.C. Chang, Auditing in the e-com era, Information Management & Computer Security 12 (5) (2004) 389–400.
[75] H. Zhu, H. Wu, Assessing the quality of large-scale data standards: a case of XBRL GAAP taxonomy, Decision Support Systems 59 (4) (2014) 351–360.

Partha Saha is a Fellow from the Indian Institute of Management Calcutta. He holds a PGDBM from the Indian Institute of Management Calcutta and a B. Tech. from the National Institute of Technology, Calicut. He is a technocrat with over 14 years of experience in IT Industry & Research. His research interests are in business intelligence, ontology based compliance, risk management, business analytics, information security etc.

Indranil Bose is a Professor and Coordinator of Management Information Systems at the Indian Institute of Management Calcutta. He is also a Coordinator of the IIMC Case Research Center and Chairperson of the program leading to the Postgraduate Diploma in Business Analytics. He holds a B. Tech. from the Indian Institute of Technology, MS from the University of Iowa, MS and Ph.D. from Purdue University. His research interests are in business analytics, information security, and supply chain management. His publications have appeared in Communications of the ACM, Communications of AIS, Computers and Operations Research, Decision Support Systems, Ergonomics, European Journal of Operational Research, IEEE IT Professional, Information & Management, International Journal of Production Economics, Journal of Information Technology Teaching Cases, Journal of Organizational Computing and Electronic Commerce, Journal of the American Society for Information Science and Technology, Operations Research Letters etc. He is listed in the International Who's Who of Professionals, Marquis Who's Who in the World, Marquis Who's Who in Asia, Marquis Who's Who in Science and Engineering, and Marquis Who's Who of Emerging Leaders. He serves as a Senior Editor of Decision Support Systems and Associate Editor of Communications of AIS, Information & Management, and several other IS journals.

Ambuj Mahanti is a Professor of Management Information Systems at the Indian Institute of Management Calcutta. He is a D.Sc. in Computer Science (University of Calcutta). He holds a post-graduate diploma in Computer Science (Indian Statistical Institute, Calcutta), a Master's in Statistics (University of Calcutta) and post-graduate diploma in Operations Research (Operational Research Society of India, Calcutta). He was a Visiting Associate Professor for nearly three years at University of Maryland College Park, USA from 1990 to 1992. His research interests include artificial intelligence, heuristic search techniques, business intelligence, cloud computing, recommender systems, ontology based compliance, risk management etc. He has wide consulting interests with a number of organizations including the Tea Board, Production Information System of Tea, Royal Norwegian Embassy, Delhi, Assessment of Training Needs for the Health Personnel in Uttar Pradesh, Hindustan Copper Limited, Identification of Computing Needs, Indian Oil Corporation, IMS: Integrated Manpower Study for the Haldia Refinery, Andrew Yule Group, CADS: Comprehensive Assessment of a Diversification Strategy for the Dishergarh Power Supply Limited etc. He has widely published in a number of journals including Computer Communications and Networks, Database and Expert Systems Applications, Theoretical Computer Science, Computational Science, network design and management, Journal of the ACM, OMEGA etc. He was a Dean (Planning and Administration) at IIM Calcutta from 2005 to 2007.