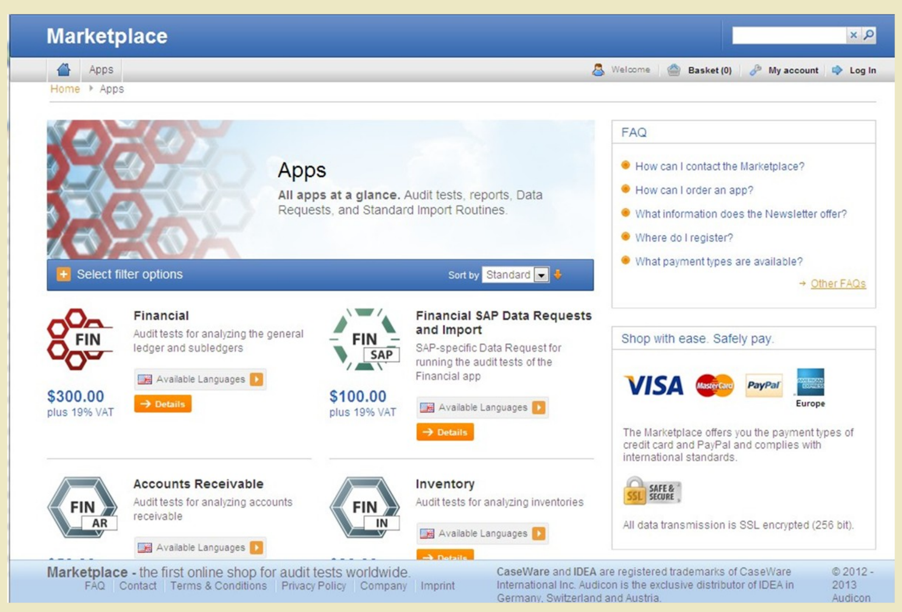
Which Audit App(s) Should Be Used? An Exploratory Study of Using Recommender Systems for Audit Apps Selection

Jun Dai, Desi Arisandi, and Miklos A. Vasarhelyi

Motivation

•Audit apps are formalized audit procedures performed through computer scripts.

Example – Caseware Marketplace



A Potential Problem

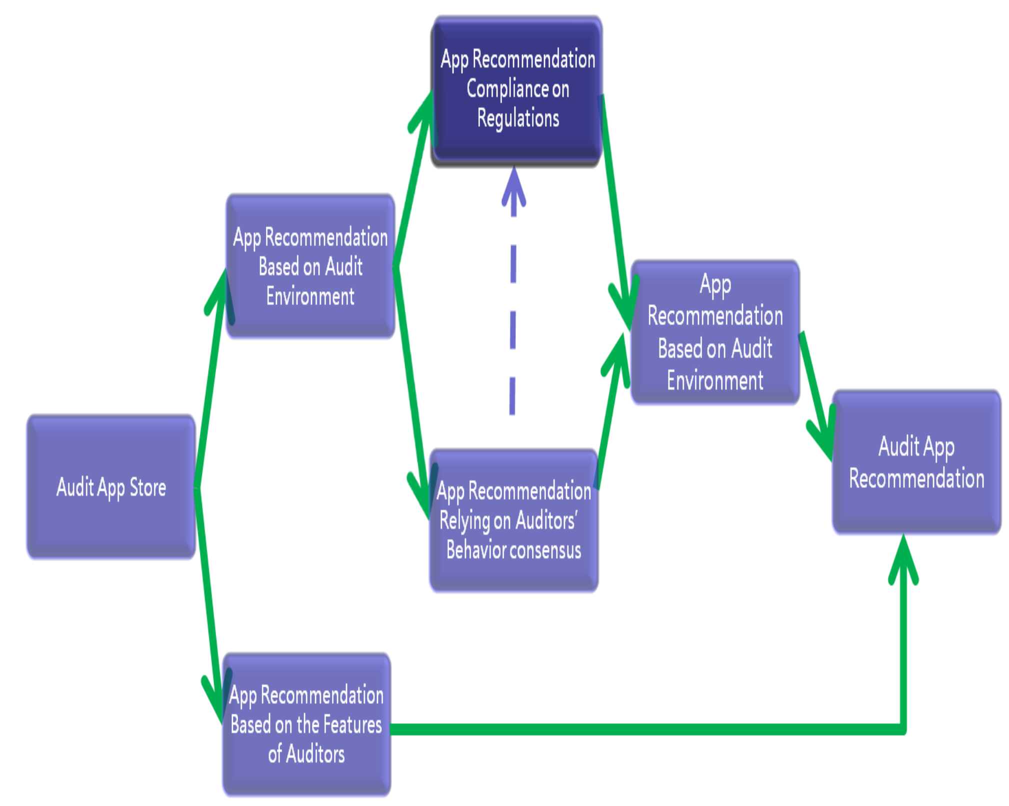
- The explosion in the number and variety of audit apps makes discovery a main challenge. This leads to the demands for recommendation systems for audit apps.

Basic Ideas

We propose a "**Super Audit App,**" which can provide guidance for auditors to ***effectively select appropriate audit apps for a particular audit client***. It is grounded on the idea of recommender systems because:

1. • Recommender systems are superior to other information filtering applications because of its ability to provide personalized recommendations.
2. • Recommender systems are able to generate results to each user that are personalized because they take each user' s personal characteristics and behaviors into account.
3. • Regarding audit app selection, recommender systems can suggest the appropriate audit apps by analyzing audit environment and auditors' historical preferences.

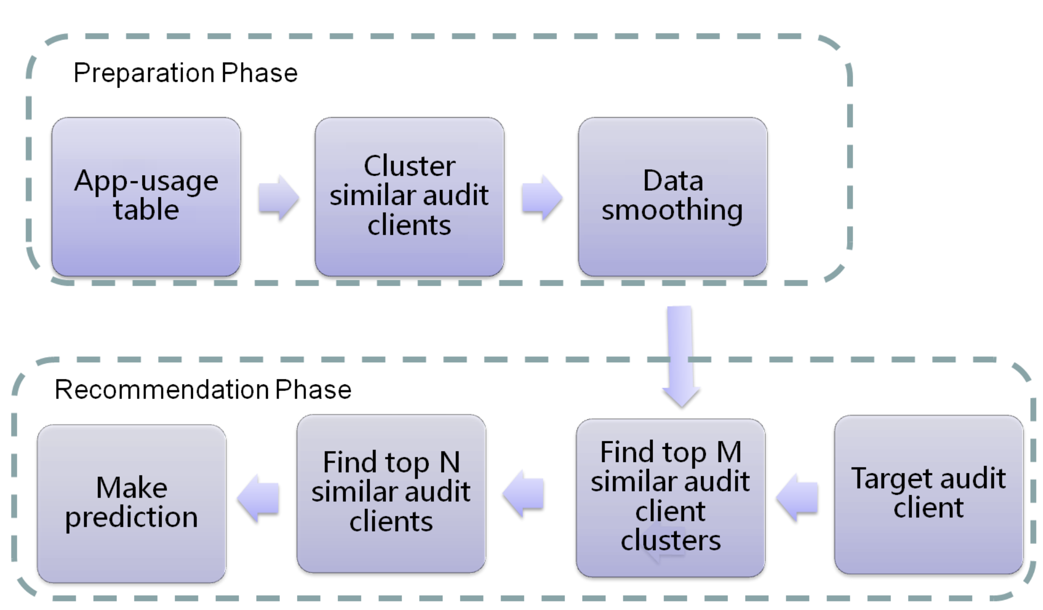
Methodology – Overview



Recommendations Based on Audit Environment

1. • Selection of Apps
2. • Business Cycle
3. • Industry
4. • Audit Assertions
5. • Audit Procedures
6. • Apps Recommendation

Recommendations Based on Audit Clients



Making a Recommendation

•Use a weighted linear model to generate a score for each audit app, combining the predicted rating from the auditor and the predicted usage for the audit client.

- *Score = δ *Rating+ (1-δ) *Usage*
- Audit apps with high scores will be recommended to the auditor in a particular audit engagement.

•Refine audit app recommendations made based on the audit environment by combining the predicted rating from the auditor .

- Audit apps with high predicted ratings will be recommended

Conclusion

- In this paper, we propose a Super Audit App framework based on recommender systems to provide digital suggestions for the auditor.
- By analyzing the audit environment and auditors' historical behaviors, the recommender systems can provide "personalized suggestions" for a particular auditor on a particular audit engagement.
- In future work, we will implement the framework and improve it according to auditor feedback.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Accounting Information Systems (AIS) in an Age of Big Data

Kevin Moffitt and Miklos Vasarhelyi

Introduction

- Accounting information systems (AIS) must accommodate business needs generated by rapid changes in technology.
- Three core assertions relative to the measurement environment in accounting, the nature of data standards for software-based accounting, and the nature of information provisioning (formatted and semantic) were discussed.
- In the area of assurance, additional concerns were cited such as traditional audit procedures hindering the performance of their objectives and the current auditing cost-benefit tradeoffs being calibrated for a different data processing era.
- The pervasive phenomenon of "Big Data" is emerging and coloring these assertions. This study aims to deal with the effect of Big Data on the issues discussed above.

What is Big Data?

Typically Big Data

- is automatically machine obtained/generated,
- may be a traditional form of data now expanded by frequent and expanded collection,
- may be an entire new source of data,
- is not formatted for easy usage,
- can be mostly useless, although Big Data is collected and its economics are positive,
- is more useful when connected to structured data in corporate enterprise systems (ERPs) (Franks, 2012 adapted).

This study shows that businesses who use Big Data to inform their decisions have **5-6% higher profitability**.

Big Textual Data

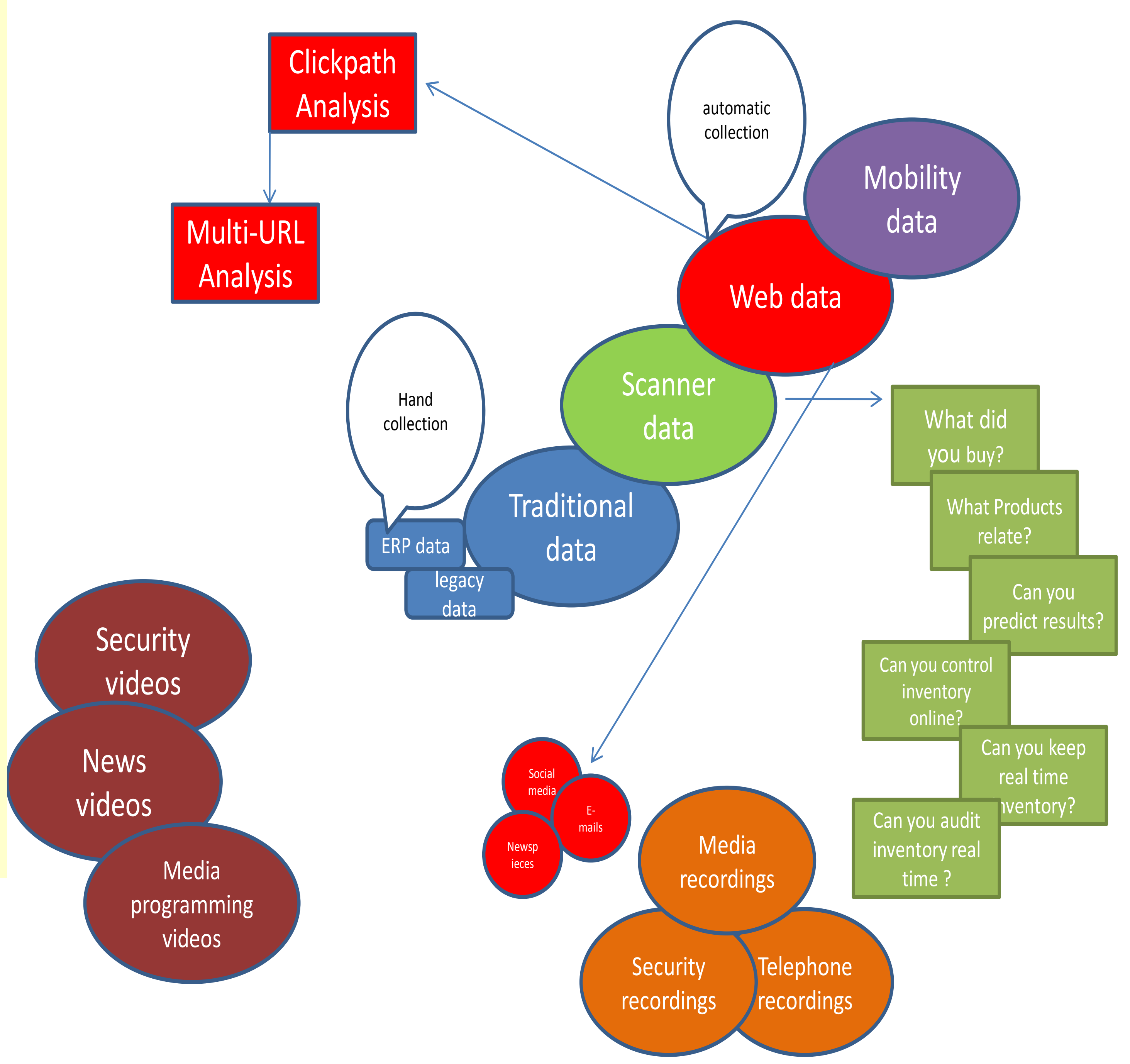
- Textual data come from many sources: Edgar, newspapers, websites, and social media.
- The text can be parsed and processed with software.
- Text understanding and vague text understanding can provide the necessary links from textual elements to the more traditional ERP data.

Big Data and Accounting Research

- There is limited accounting research that uses Big Data to derive results.
- With open source software and commoditized hardware, Big Data should be available for accounting research.
- The challenge is for accounting researchers to become data intensive when organizational data is not always easy to obtain.
- In addition, most data is not standardized and there may be substantive cost in its pre-processing.

Big Data in Business Management, Assurance, and Standard Setting

The Enterprise Data Ecosystem (EDE) is exponentially expanding:



Data Expansion and Measurement Consequences

Sources, content and enhanced content (content)	Parameters (meta data)	Meta-meta data	New Sources of Business Data	New Sources of financial data	New Linkages facilitated by IT and Analytic Technologies
News-pieces			RFID data		
	Source		Detailed transaction data		Use news mentions or e-mails or click path analysis to predict sales
	Time		B2B transaction data	B2B transaction data	
	Publication		Blog postings		
	Topic	Frequency over time	e-mails		
		Change in nature over time	Social media postings		
e-mails			e-mails linked to transactions with attachments	XBRL/FR data	Automatic classification and addressing of e-mails
	To whom			XBRL/GL data	
	From whom			Edgar data	
	Topic			FDIC call reports	
		Frequency over time			
		Change in nature over time			
Social Media		Groups of people that behave similarly		Comments on blogs about companies	Automatic response and tone detection
	To whom				
	From whom				
	topic				
Click-path		Purchase paths by product		Paths of fraudulent behavior	Identify fraudulent user groups
	Website visited				
	Pages visited				

Conclusion

- The accounting model must evolve/be changed to focus on data content, atomicity, data linkages, etc.
- Accounting standards will have to deal with the large databases and allowable sets of extractions, not with extant rules of disclosure.
- Automatic confirmation will limit the need for verification of population and data integrity.
- Auditors should seek to verify transactions not with just an invoice and receipt, but multi-modal evidence that a transaction took place.
- Public good would be served if large research-oriented public financial related databases could be made available to the accounting research community.
- Accounting education will have to evolve educating faculty, professionals, and students in the issues of Big Data and data analytics.

Exploratory Visual Analysis of Medicare Health Insurance Data for the Purpose of Knowledge Discovery

Abdullah Alawadhi

Paul Byrnes

Introduction

King (2010) views Medicare as a high-risk program, at least partially because of its size and its complexity, which creates opportunities for abuse, waste, and fraud.

Additionally, the vast amount of data generated by healthcare insurance providers today is far too complex and voluminous to be processed and analyzed via traditional methods.

Consequently, providers must rely on advanced techniques to find and track offenders (Koh & Tan, 2011).

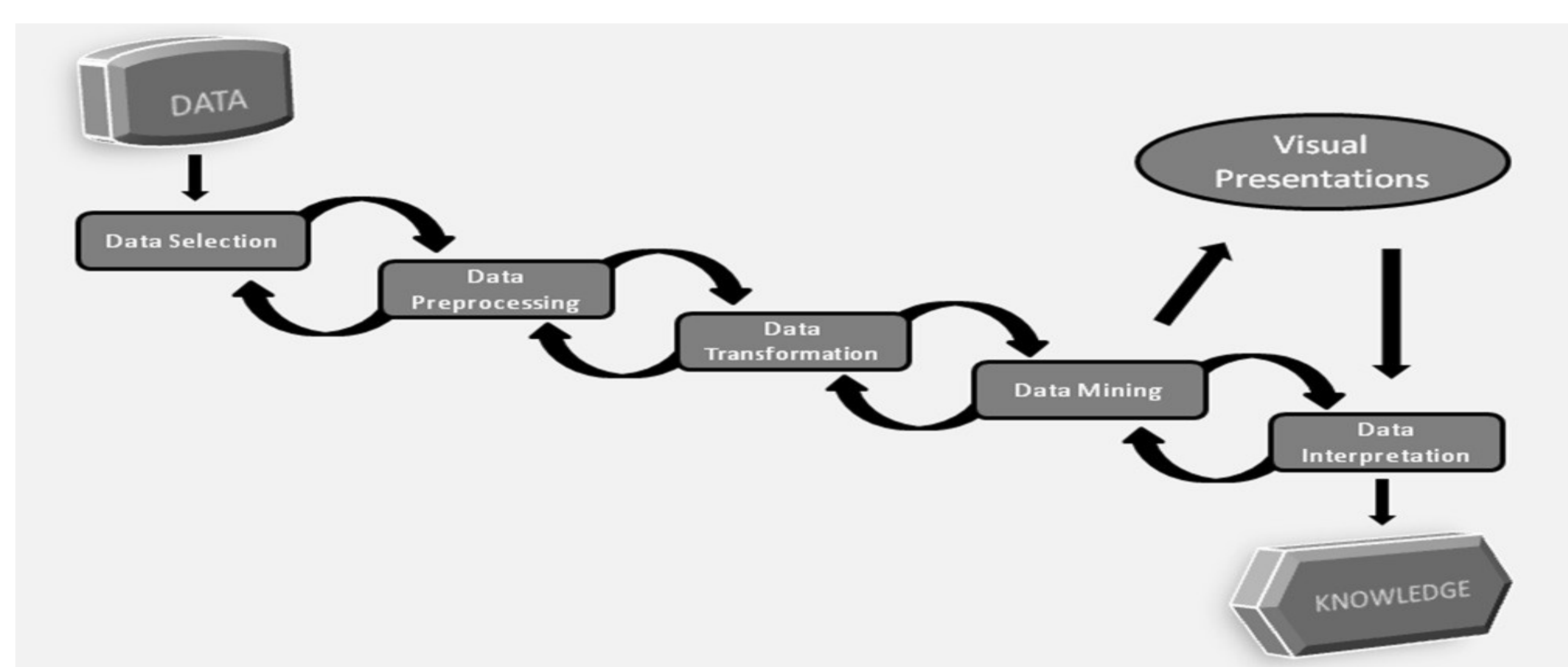
Data and Methodology

The data:

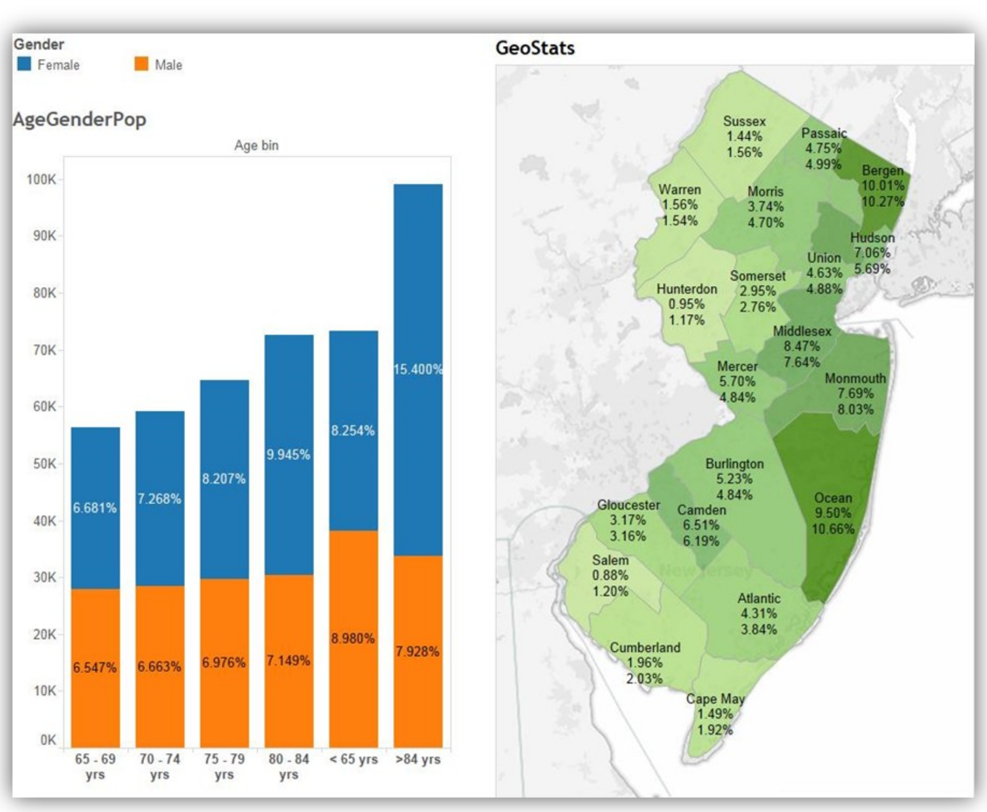
Medicare patient claims and provider details for the 2010 fiscal year for all 50 states in the U.S., with about 12 million records and more than 1,600 attributes.

Exploratory Visual Analysis (EVA):

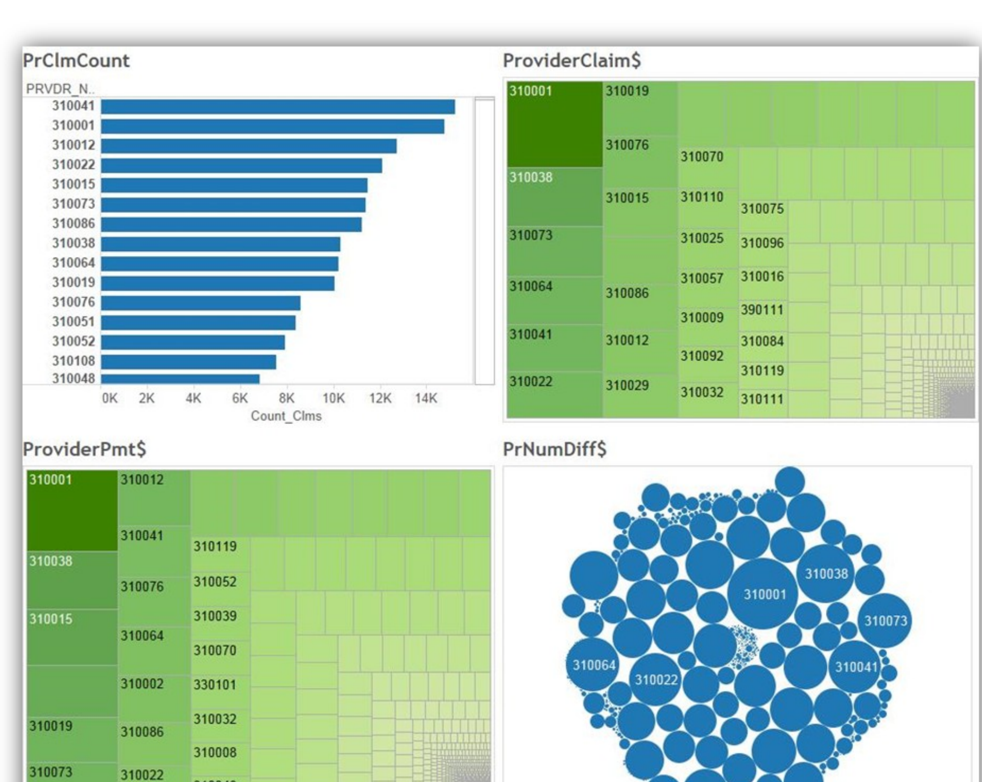
EVA is technique that accommodates large datasets and can assist in the process of knowledge discovery which ultimately helps find useful and valid information in large volumes of data. One advantage of this method involves user interactivity, whereby a human is actively engaged in the data exploration process, leveraging his/her abilities to perceive patterns and structures in visual representations and ultimately interpret what is seen.



Results and Discussion



County	Population	Percent of Total Population
Atlantic	274,549	3.12%
Bergen	905,116	10.29%
Burlington	448,734	5.10%
Camden	513,657	5.84%
Cape May	97,265	1.11%
Cumberland	156,898	1.78%
Essex	783,969	8.92%
Gloucester	288,288	3.28%
Hudson	634,266	7.21%
Hunterdon	128,349	1.46%
Mercer	366,513	4.17%
Middlesex	809,858	9.21%
Monmouth	630,380	7.17%
Morris	492,276	5.60%
Ocean	576,567	6.56%
Passaic	501,226	5.70%
Salem	66,083	0.75%
Somerset	323,444	3.68%
Sussex	149,265	1.70%
Union	536,499	6.10%
Warren	108,692	1.24%
Total	8,791,894	100.00%



Utilizing EVA, we provide tabular descriptive, graphical representations, and visual dashboards. By merely glancing at the visualization dashboard, questions immediately surface and the outliers become glaringly apparent such that investigatory resources might be better allocated for resolution purposes.

Contribution and Future Work

Limited research has been carried out concerning the uses of EVA as the primary enabling technology for comprehensive knowledge construction. Hence, the aim and contribution of our study is to help bridge this gap by relying on EVA for knowledge discovery.

This project was very beneficial and allowed us to obtain some degree of insight about visualizing data for the purposes of information extraction and knowledge discovery. In the future, perhaps we will be able to further assist in the Medicare review and recovery process by continuing knowledge discovery work in this domain.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Expert Knowledge Elicitations in a Procurement Card Context: Towards Continuous Monitoring and Assurance

Abdullah Alawadhi

Deniz Appelbaum

Introduction

Employee procurement cards have been commended by public and private companies for their success in reducing purchasing department costs and increasing individual department purchasing decisions (Daly and Buehner, 2003).

However, recently critical internal controls are lacking in the area of procurement cards and the likelihood of employee fraud has drastically increased (Gillett, 1997).

Furthermore, advances in data processing, information technology, as well as the rise of ERP systems, have facilitated the creation of real-time accounting information. Thus, a Continuous Auditing (CA) system is necessary to help auditors and management provide continuous monitoring and assurance of internal controls and detect any exceptions on a timely basis by relying on technology throughout the audit process (Alles et al. 2006).

Motivation

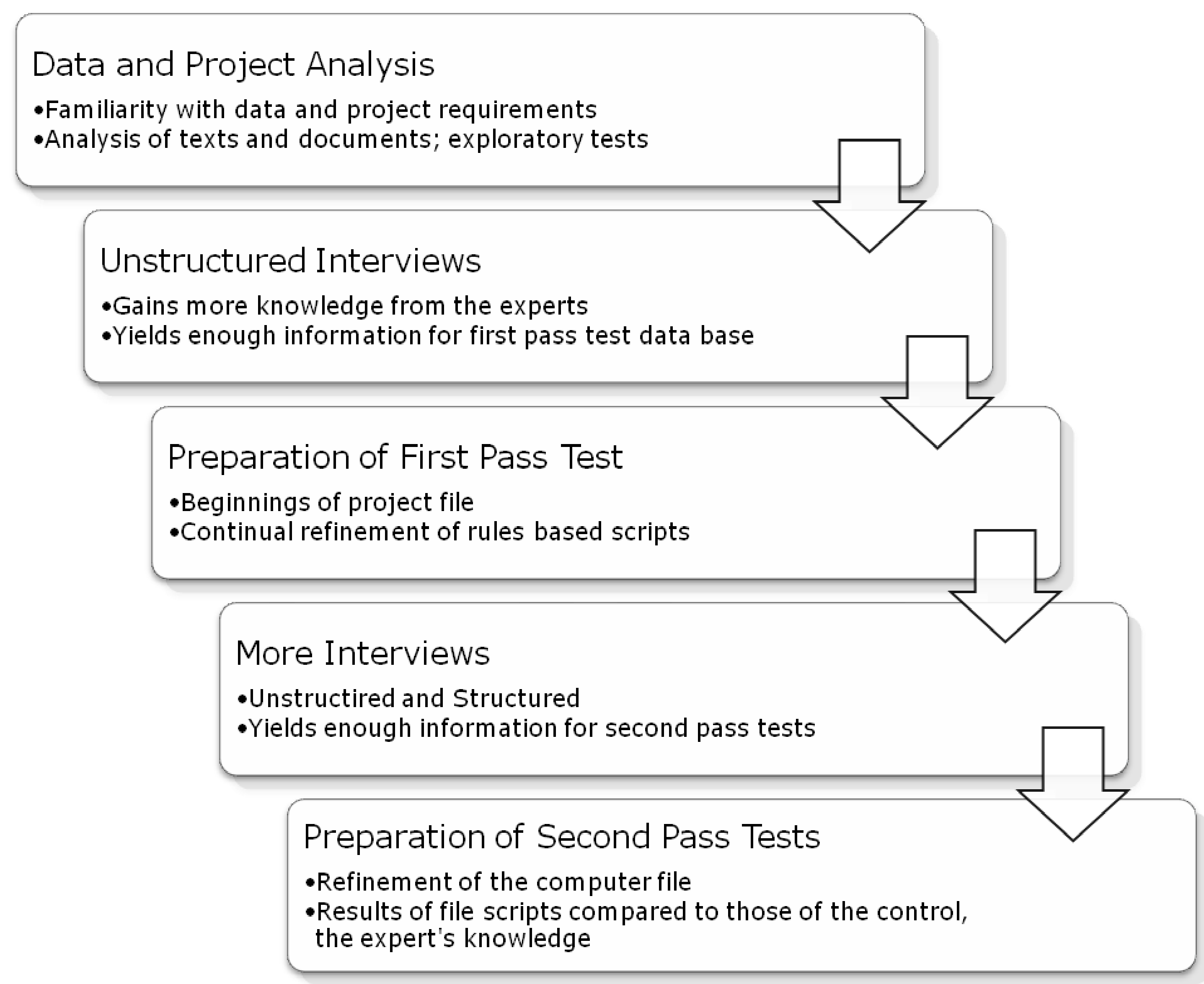
There are many studies of general fraud and credit card fraud (Quah and Sriganesh, 2007), but few could be found evaluating procurement card issues within a corporate setting. This lack of research attention is primarily due to the scarcity of feasible data (Amat, 2002).

Therefore, the contribution of this study is to bridge this gap by utilizing real world procurement data and applying a multi-dimensional approach for p-card misuse detection

Data and Methodology

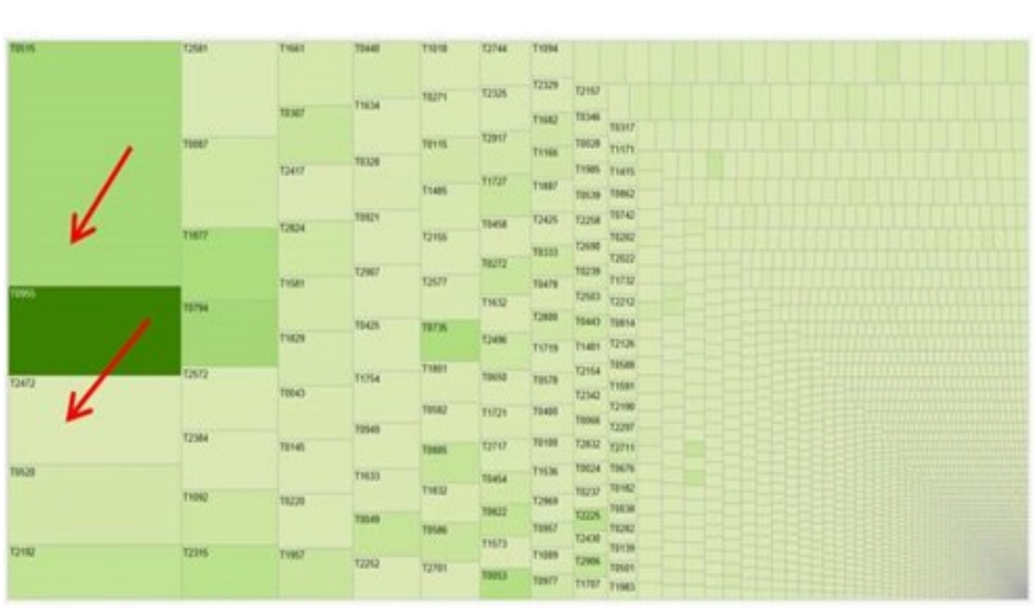
The data covers all of 2011 and 2012 and is updated monthly at month-end, averaging 50,000 transactions per months with 51 total attributes.

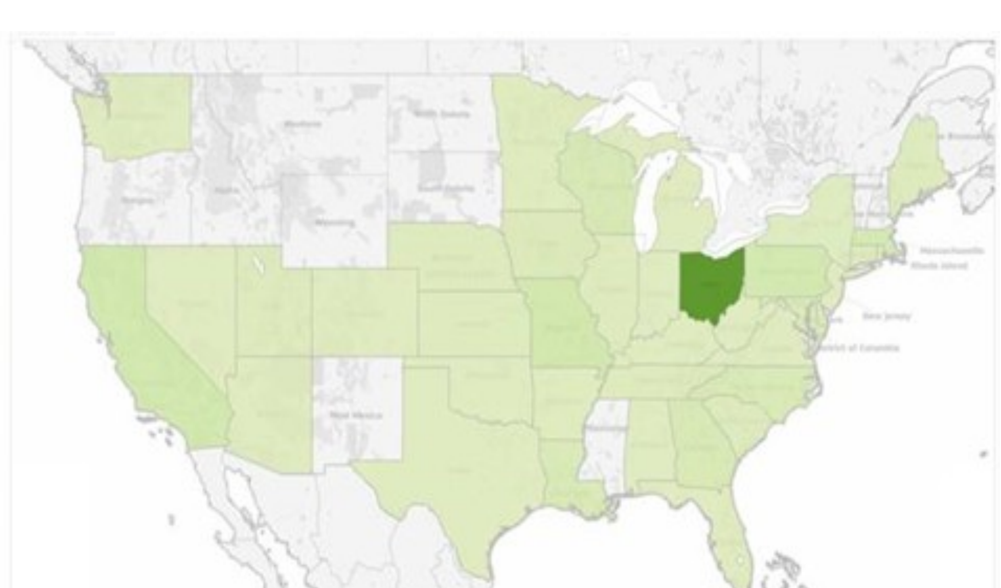
The purpose of this project is to develop an expert system from a domain expert's knowledge. The entire project structure is shown below:



Results

Our initial run of the expert system produced a total of 1408 exceptions. After reviewing the exceptions with the experts, 68% of the pulled exceptions were considered legitimate red flags and would require further investigation. For the remaining 32% of exceptions, new rules will be formulated with the experts. Additionally we plan on building user profiles and utilizing visualization scenarios to further assist in outlier detection





ID	Purchase Date	Original Currency Amount	Extended Item Amount	Merchant Name	Item Description
ID1637	2/17/2011	0	50	STAPLES 00101907	$ 50 APPLES ITUNES
ID1917	2/22/2012	0	7.59	TARGET 00014472	POKER CHIPS 11.5G GAME ESSEN
ID0925	3/25/2011	84.95	75	AMAZON MKTPLACE PMTS	ITUNES GIFT CARD
ID4720	7/22/2011	0	10	BOLDEN INSTRUMENT	FUEL CHARGE $10
ID2503	10/6/2011	31.95	31.95	AMAZON MKTPLACE PMTS	PROACTIV SOLUTION ORIGINAL REPAIRIN
ID0305	10/11/2011	16.28	12.99	AMAZON.COM	CONAIR TOUCH AND TONE MASSAGER WITH
ID2315	10/11/2012	49.69	41.66	STAPLES	STRESS BUSTER MASSAGE FOOT
ID5477	11/14/2012	24.5	22	AMAZON MKTPLACE PMTS	BRIDAL WEDDING JEWELRY HAIR HEADBAN

Conclusion

The project is still a work in process, primarily due to the complexity of rules and transactions that must be gleaned in this outlier detection.

In the future, the expectation is that this multi-pronged simultaneous attack of using automated rules and generated profiles to detect unusual p-card activity will assist enormously in the arena of occupational fraud detection.

Furthermore, these scripts can be applied on a continual basis, contributing to the continual journey of expert knowledge elicitation in a continuous auditing and monitoring environment.

RUTGERS
Rutgers Business School
Newark and New Brunswick

Using XBRL to Conduct a Large-scale Study of Discrepancies between the Accounting Numbers in Compustat and SEC 10-K Filings

Roman Chychyla and Alexander Kogan

Introduction

The Compustat accounting database is frequently used for both research and decision-making.

It has been documented (San Miguel 1977; Rosenberg and Houglet 1974; Yang et al. 2003; Tallapally et al. 2011, 2012; Boritz and No 2013) that information found in the Compustat database differs from both the information found in other accounting databases and the information disclosed in corporate financial filings.

In this study, we conduct the first large-scale comparison of Compustat and 10-K data. Specifically, we compare 30 accounting items for approximately 5,000 companies for the period from October 1, 2011 to September 30, 2012.

We utilize the power of the XBRL reporting technology to automatically extract accounting numbers from XBRL 10-K filings and compare them to Compustat numbers.

Discrepancy Examples

In its 2011 10-K statement, Amazon.com reported Cost of Sales to be $37,288M. Compustat adjusted this number to $36,288M by excluding $1B of Depreciation Expense.

Transwitch Corp's 2011 Gross Profit of $19,932,000 was reported in Compustat as $17,932,000 (the difference of $2M (10%) is probably due to input typo).

ADA-ES restated its Net Loss from $19.851M (March 12, 2012) to $22.819M (October 19, 2012). As of February 22, 2013, Compustat did not update the value resulting in a difference of $2.968M (13%).

Compustat did not report the 2012 Total Assets of Airwave Labs ($1,139,182), although Compustat did report the company's Total Liabilities and Stockholder's Equity.

Types of Discrepancies

There are four main reasons for data differences between Compustat and 10-K reports:

1. Compustat adjusts company reported numbers to match its standard definitions of variables "for ensuring. . . comparable data across companies, industries and time periods without reporting biases or data discrepancies" (Compustat's website).
2. Compustat's value is erroneous (due to typos, rounding, etc.).
3. Compustat's value is not up to date.
4. Compustat does not provide a value for the data item.

Methodology

We develop a methodology to cross-verify Compustat data using XBRL 10-K reports that consists of the following six steps:

1. Extracting data from Compustat.
2. Extracting data from XBRL 10-K filings.
3. Merging Compustat and XBRL data.
4. Creating mappings between Compustat variables and XBRL reporting concepts.
5. Calculating differences between Compustat variables and the associated XBRL reporting concepts.
6. Analyzing discrepancies between Compustat and XBRL 10-K filings.

Removing XBRL Errors

XBRL 10-K reports may (and do) contain errors. Since we want to compare Compustat to (plain-text) 10-K data, it is necessary to remove erroneous XBRL observations from the matched sample.

We developed automated procedures to

1. find specific set of XBRL errors and idiosyncrasies (e.g., wrong sign, wrong scale, use of extension, wrong tag name, non-conventional dimension, etc.) and
2. reconcile Compustat and XBRL discrepancies.

We manually check unexplained discrepancies and "suspicious" reconciliations to determine whether they are XBRL-related errors. We have manually cross-verified 1,800 discrepancy items (around 1.5% of all items) using original 10-K filings.

Results

We find that:

- Compustat values of 22 out of 30 analyzed variables significantly differ from values reported in the 10-K filings.
- Variables with fairly simple definitions (e.g., Total Assets = All Assets, Total Liabilities = All Liabilities, Net Income = All Revenues - All Expenses) tend to have less discrepancies than variables that have more complex definitions (e.g. Cost of Goods Sold).
- The type of statement where variables are reported and company characteristics such as industry and size are related to the amount and magnitude of discrepancies.

RUTGERS
Rutgers Business School
Newark and New Brunswick

The Application of Exploratory Data Analysis (EDA) in Auditing

Qi Liu and Miklos Vasarhelyi

Motivation

- Auditing is a data intensive process; data analysis plays an important role in the audit process.
- Current data analysis approaches used in the auditing process focus on validating predefined audit objectives, which are unable to discover unaware risks from the data.
- EDA is often linked to detective work and one of its objectives is to identify outliers.
- Even though some EDA techniques have been used in some auditing procedures, EDA has never been systematically employed in auditing.

Definition of EDA

- Exploratory data analysis (EDA) is a data analysis approach emphasizing pattern recognition and hypothesis generation.

Application of EDA Techniques in Auditing

- **Analytical Review**
  - Descriptive Statistics
  - Data Visualization
  - Data transformation
- **Controls Assessment and Testing**
  - Process Mining
- **Substantive Testing of transactions and details of balances**
  - Data transformation
  - Feature Selection
- **Fraud Detection**
  - Descriptive Statistics
  - Data Visualization
  - Cluster Analysis
  - Association Analysis
  - Social Network Analysis

Potential Applicable Areas in Audit Standards

AU-C 240	*•Performing a detailed review of the entity's quarter-end or year- end adjusting entries and investigating any that appear to have* ***an unusual nature or amount*** *•Analyzing sales discounts and returns for* ***unusual patterns or trends*** *•Reviewing the propriety of* ***large and unusual expenses***
AU-C 315	*•Analytical procedures performed as risk assessment procedures may identify as-*
AU-C 520	*•The results of analytical procedures designed and performed near the end of the audit may identify a* ***previously unrecognized*** *risk of material misstatement.*
AU-C 550	*•...****an unusually high turnover*** *of senior management or professional advisors may suggest unethical or fraudulent business practices that serve the related party's purposes.* *•In evaluating the business rationale of a significant related party transaction… the auditor may consider… Whether the transaction (1)* ***has unusual terms*** *of trade, such as unusual prices, interest rates, guarantees, and repayment terms (2) is processed in* ***an unusual manner***

Application of EDA in the Audit Cycle

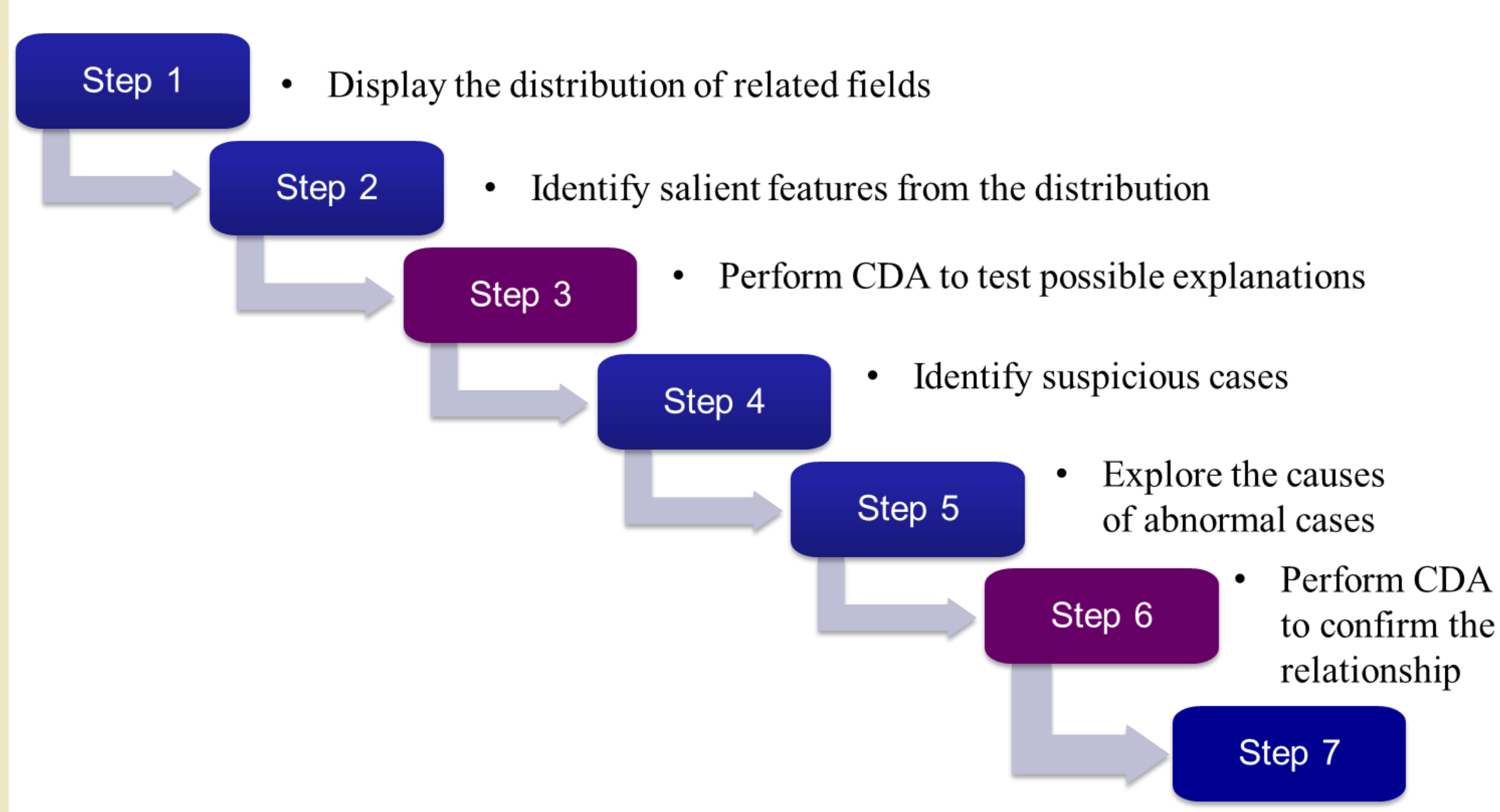
- Assess and respond to engagement risk
- Understand client's business
- Assess client business risk
- Perform preliminary analytical procedures

- Assess acceptable audit risk and inherent risk
- Understand internal control and assess control risk
- Assess fraud risks

- Assess engagement quality
- Include EDA results in audit report

Preprocess data to facilitate auditors to perform analytical procedures

Framework to Apply EDA in Auditing



- Report the risks and recommend improvement suggestions
- Add a new audit objective

RUTGERS
Rutgers Business School
Newark and New Brunswick

Customer Segmentation Via Clustering: A Two Stage Approach

Paul Byrnes

Abstract

In this paper, the popular K-means algorithm is used in segregating credit card customers of a large banking institution located in South America. Interestingly, it is found that a two-stage approach to the clustering exercise seems to best segment the existing customer data. In the first stage, the K-means method is executed on the data set in an effort to identify potentially optimal solutions. Following this, the DBScan algorithm is initiated in order to determine whether support for any of the selected K-means models exists relative to both number of clusters as well as composition. In fact, this procedure leads to the preliminary identification of a five cluster K-means solution. In the second stage, each of the five clusters is analyzed separately using a comprehensive set of mechanisms and procedures in order to determine whether distinct and multiple sub-clusters are contained within a given primary cluster. Findings in the second phase clearly suggest that three of the primary clusters each contain two sub-clusters. Consequently, while the first stage provides evidence for only five customer segments, the second stage argues for the existence of eight. In conclusion, it is determined that eight clusters best distinguishes the analyzed data.

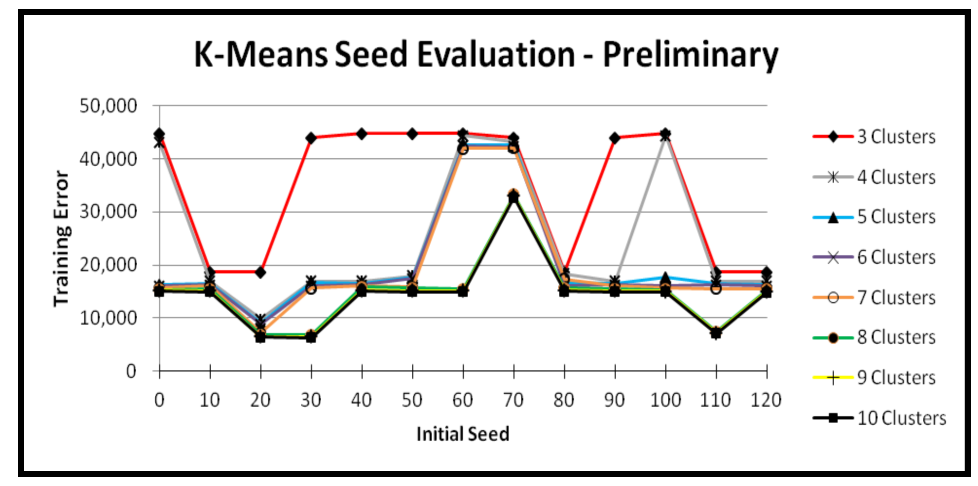
Introduction

Fayyad et al. (1996) note that clustering is a well-established approach for finding worthwhile patterns in data. Furthermore, Tan et al. (2006) indicate that clustering has been effectively employed to address an extensive array of issues, including customer segmentation activities. In a generic sense, cluster analysis entails placing data into groupings that are meaningful and useful, such that each object is comparable to items in the same cluster and different from objects assigned to other clusters.

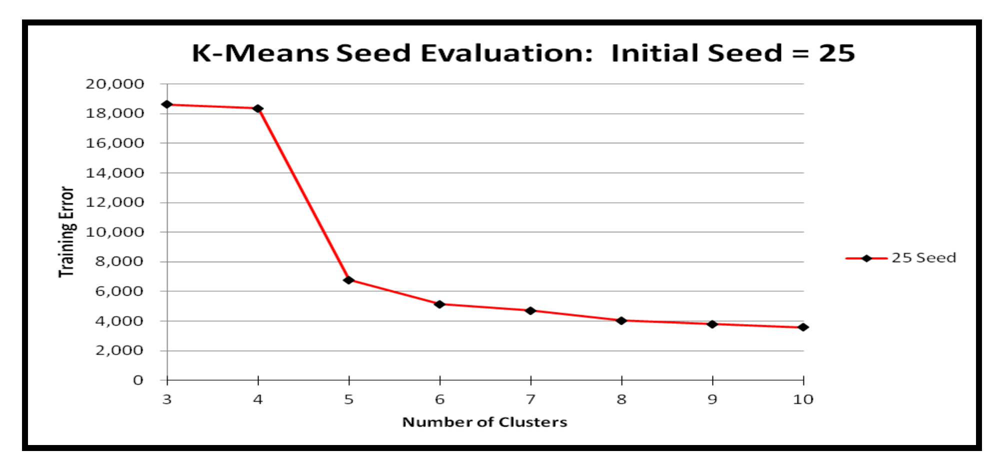
In this study, the K-means algorithm is primarily relied upon as a method for segregating the customer base of a large financial firm based upon an extensive, preprocessed data set containing five key dimensions and 186,722 records. To assist with this activity, a variety of supplemental mechanisms are employed including DBScan, silhouette evaluation, elbow analysis, descriptive statistics, and statistical testing. The results of a two-stage procedure suggest that eight groups are needed to fully segment the customer base.

Analysis-Stage 1

Initially, seed values between zero and 120, in increments of 10, are explored. Furthermore, each of these 13 initial seeds is paired with an array of cluster numbers ranging from three to 10, inclusive. Subsequently, model results are evaluated using training error (SSE) as the criterion for model assessment.



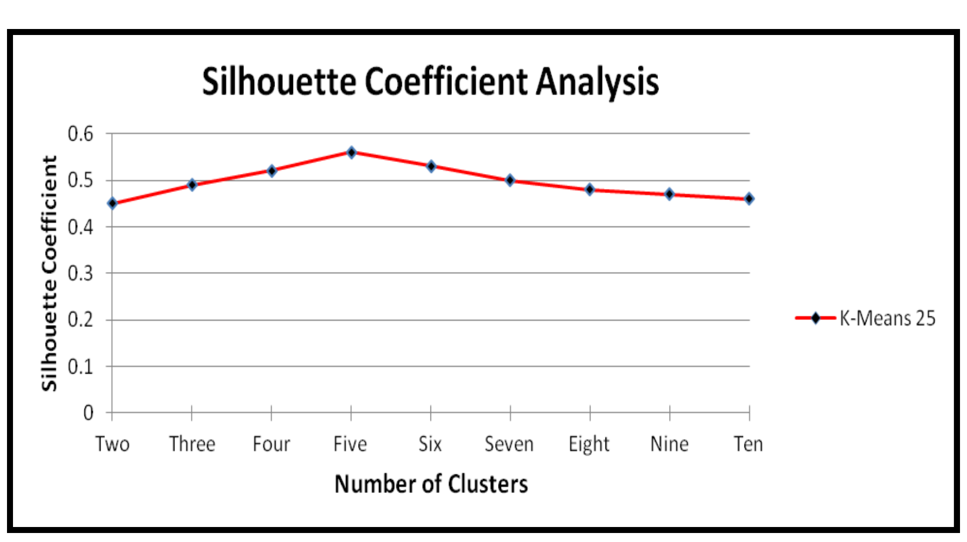
It is noted that relatively low error occurs at initial seed amounts of 20, 30, and 110. Given this, all possible seed values between 15 and 35, and 105 and 115 are paired with cluster numbers ranging from three to 10 so as to construct a more detailed set of K-means solutions. Upon completion of this exercise, it is found that initial seedings of 19, 20, 23, 25, and 113 seem to produce the best overall clustering results. Following is a graph of the most preferred seeding in terms of training error for all relevant models. With this seeding, it might be argued that 5 clusters exist.



Analysis—Stage 1 (cont.)

After K-means model building and analysis, DBScan is executed to further assist with determining the number of clusters parameter. In fact, it is found that DBScan proposes a five cluster model, and this corresponds with the K-means 25 seed result. Furthermore, the cluster composition of both models is found to be identical.

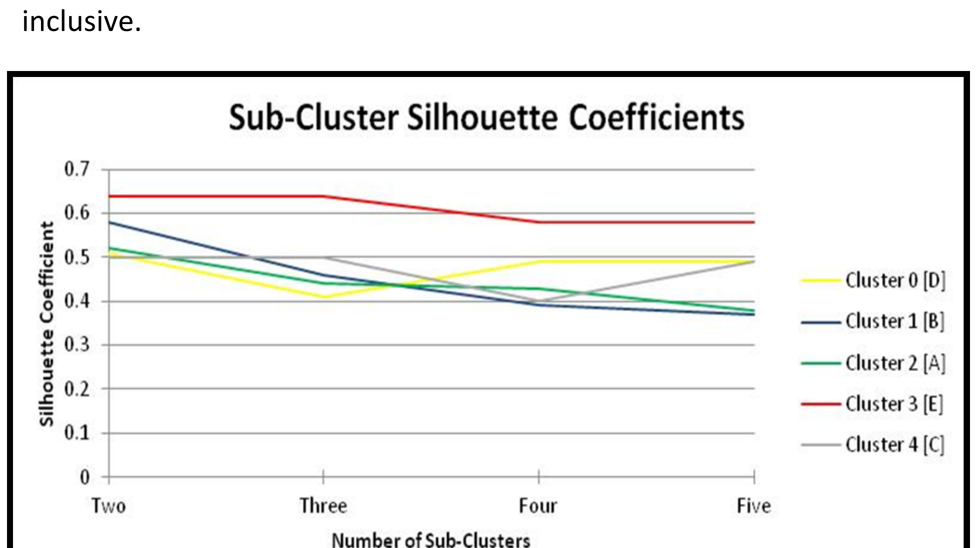
In a final confirmation procedure, a series of silhouette coefficients are computed for the selected K-means solution by varying the number of clusters from two to ten, inclusive. Incidentally, the silhouette coefficient is a measure of cluster cohesion and separation, and, thus, can provide utility in deciding upon the appropriate number of clusters. Basically, the silhouette coefficient can vary between –1 and 1, and higher coefficients are indicative of better cohesion and separation. Results of this analysis follow, and indeed suggest that, among the eligible K-means 25 seed models, five clusters offer the best performance.



At this point, all indications are that five groups most adequately segregate the evaluated data. Nevertheless, each cluster is evaluated individually in a secondary stage to increase confidence level of the findings.

Analysis—Stage Two

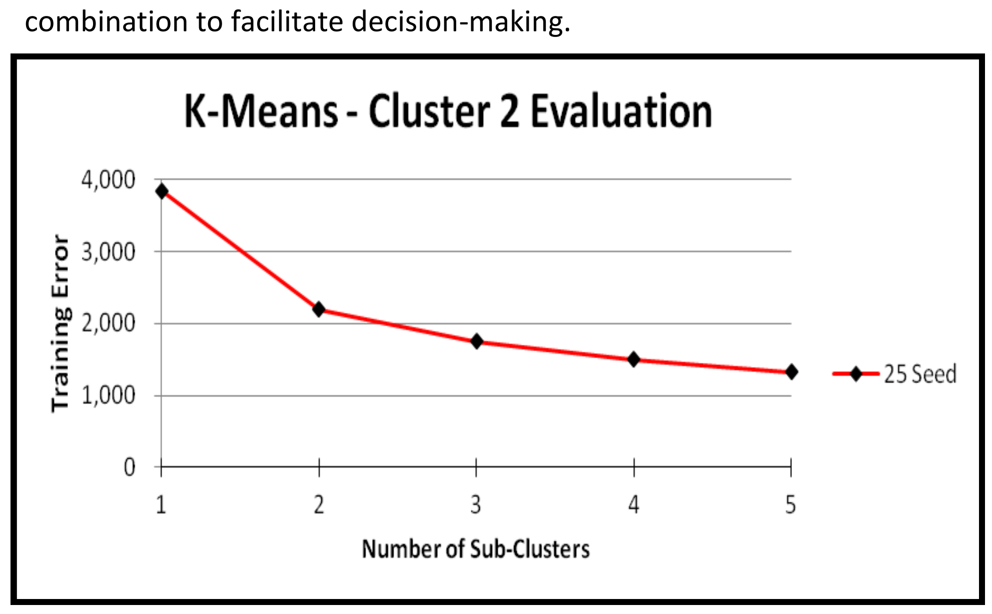
Each data subset is clustered separately via K-means to determine whether multiple sub-clusters might exist within the primary groupings identified in stage one of analysis. To assist in the evaluation process, silhouette coefficients are first computed for each of the primary clusters, whereby the number of sub-clusters is varied between two and five, inclusive.



Analysis—Stage 2 (cont.)

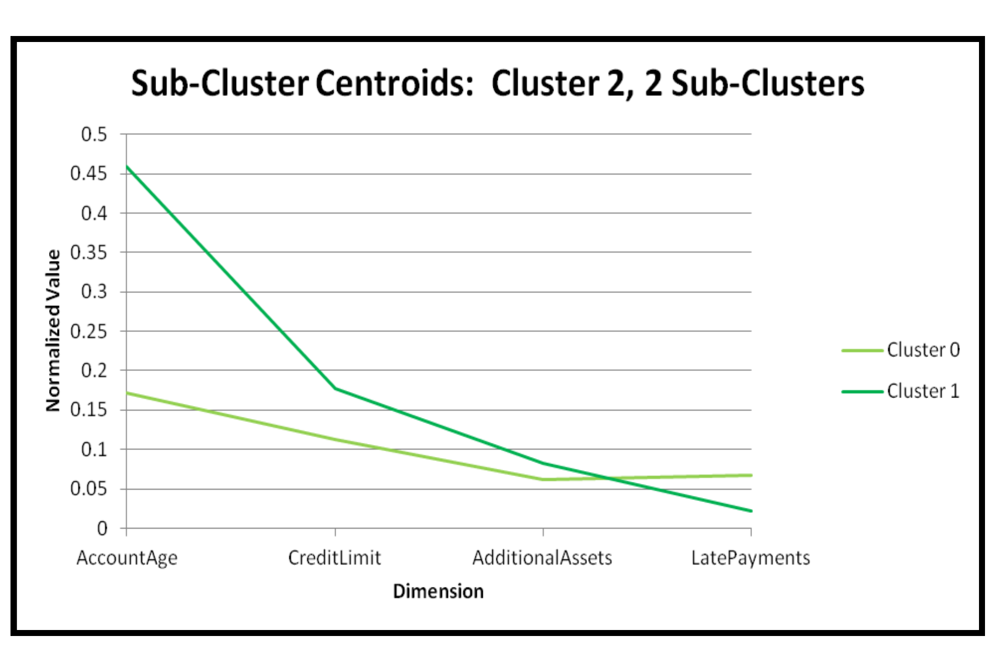
In reflecting upon the silhouette coefficients, two immediate observations surface. First, there is a general trend of decline in coefficient values as the number of sub-clusters increases from two to five. Second, the only number of sub-clusters for which all silhouette coefficients would likely be considered as good is two. While this procedure does not confirm that multiple sub-clusters exist, it provides evidence that, at most, two sub-clusters are contained within any given primary cluster.

Moving forward, SSE and elbow analyses are once again used in combination to facilitate decision-making.



In the above diagram, an elbow clearly appears at two sub-clusters, demonstrating that significant diminishing returns in error reduction occur in moving from two to three sub-clusters. This suggests that two groups might exist within primary cluster 2 of stage one. Comparable evaluations are performed for the remaining four clusters, and findings suggest that two sub-clusters are contained within three of the original clusters.

At this juncture, a K-means model is generated for each of the data subsets, such that number of clusters is designated as two. In addition, statistical testing of centroid values is conducted.



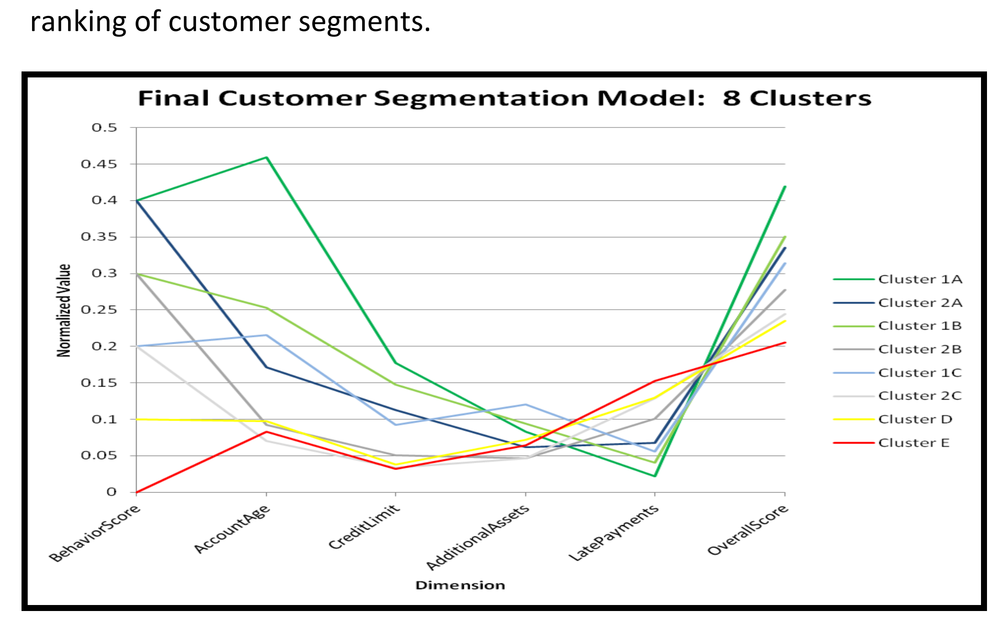
Visual inspection of the above figure suggests the existence of two subgroups. Specifically, in contrast to cluster 0, cluster 1 relates to relatively more mature accounts, whereby customers have higher credit limits, carry more additional assets with the bank, and make fewer late payments. Furthermore, an F-test of centroid values rejects the null

Analysis— Stage 2 (cont.)

hypothesis of identical means for the two sub-clusters ($p<.05$), providing additional confidence that two groups exist in this case. Comparable analyses are performed for the remaining four clusters, and results for three of the five models ague for the presence of two sub-clusters. Consequently, based upon the data examined, it is determined that eight clusters are needed to fully segment the customer base.

	Cluster							
Dimension	**1A**	**2A**	**1B**	**2B**	**1C**	**2C**	**D**	**E**
BehaviorScore	A	A	B	B	C	C	D	E
AccountAge	0.4593	0.1717	0.2529	0.0923	0.2153	0.0703	0.0971	0.0831
CreditLimit	0.1774	0.1123	0.1478	0.0507	0.0922	0.0340	0.0377	0.0323
AdditionalAssets	0.0827	0.0614	0.0943	0.0464	0.1203	0.0461	0.0718	0.0644
LatePayments	0.0216	0.0674	0.0406	0.1004	0.0561	0.1288	0.1295	0.1527
Instances:	28,422	51,840	25,147	42,495	7,980	18,976	8,858	3,004

In a final procedure, an attempt is made to depict the final eight clusters visually in terms of normalized values. Furthermore, a rudimentary scoring formula is developed and incorporated in an effort to better facilitate ranking of customer segments.



Analysis Tools: SAS, SPSS, R-Studio, and Weka

Rutgers Business School
Newark and New Brunswick

Audit Ecosystem Proposal: Definitions, Attributes, and Agents

Stephen Kozlowski

Introduction

The purpose of this research is to define an audit ecosystem, that is, the environment in which computer-based continuous auditing and monitoring (CA/CM) tools can operate with the greatest efficiency and effectiveness in order to provide the greatest benefit to both client and provider. The development of an audit ecosystem is the natural progression in the deployment of computer-based CA/CM tools, and as with earlier CA development efforts, this activity is preferably undertaken in the academic community.

Literature Review

The starting point in this definition of an audit ecosystem begins with a review of both current and significant articles in the areas of robotics, digital ecosystems, and software agents, from which the information incorporated into the audit ecosystem proposal is based.

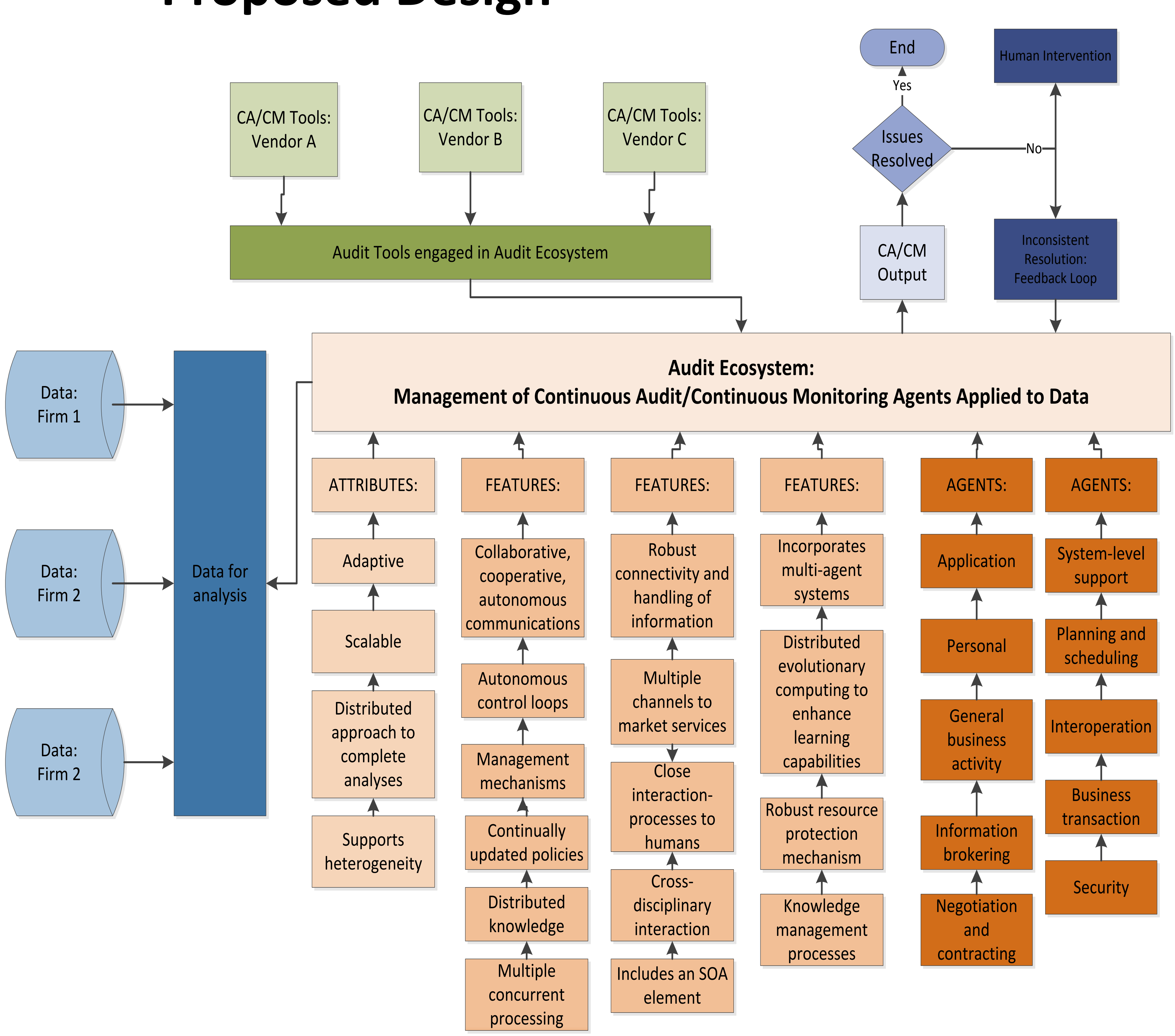
The basis for practical robotics dates back to 1948 with Norbert Wiener and development progressed with the introduction of programmable robots in the 1950's and mobile robots in the 1960's, and continues today with an increasing proliferation of robots being deployed.

Literature Review

The concept of a digital ecosystem originated in the early part of the 21st century, triggered by the European Commission-sponsored Go Digital initiative, whose aim was to boost the adoption of information and communication technologies (ICT) by European small and medium -sized enterprises (SMEs) .

Software-based agent research is generally accredited with beginning in the 1980's. The goal in the development of agent-based software was to create software with the ability to interoperate, with other programs.

Proposed Design



Software Agent Deployment within the Audit Ecosystem

Process	Agent
Prepare auditee profile	Personal agent
	Security agent
	Negotiation/Contracting agent to buy services
	System-level support agent to support transaction processing
Prepare auditor profile	Personal agent
	Security agent
	Negotiation/Contracting agent to sell services
	System-level support agent to support transaction processing
Audit tool selection/Recommender system	Application agent
	General business agent
	Business transaction agent
	Negotiation/Contracting agent to buy tool/services
	Security agent
	System-level support agent to support transaction processing
Import auditee data into ADS format	Application agent
	Security agent
Select appropriate audit tools to execute	Application agent
Capture audit rules	General business agent
Apply rules to audit tools	General business agent
Submit audit tools to execute	Application agent
	Interoperability agent to coordinate tools from different providers
Coordinate concurrent audit test processing against big data	Planning and scheduling agents
Record results of testing	Application agent
Record testing anomalies that require further research	Application agent
Feedback loop for unresolved or inconsistent results; search for resolutions	Application agent
	Information brokering agent
Unresolved issues for human review and intervention	Personal agent
	Security agent
	Application agent
Update audit rules based on testing	Information brokering agent

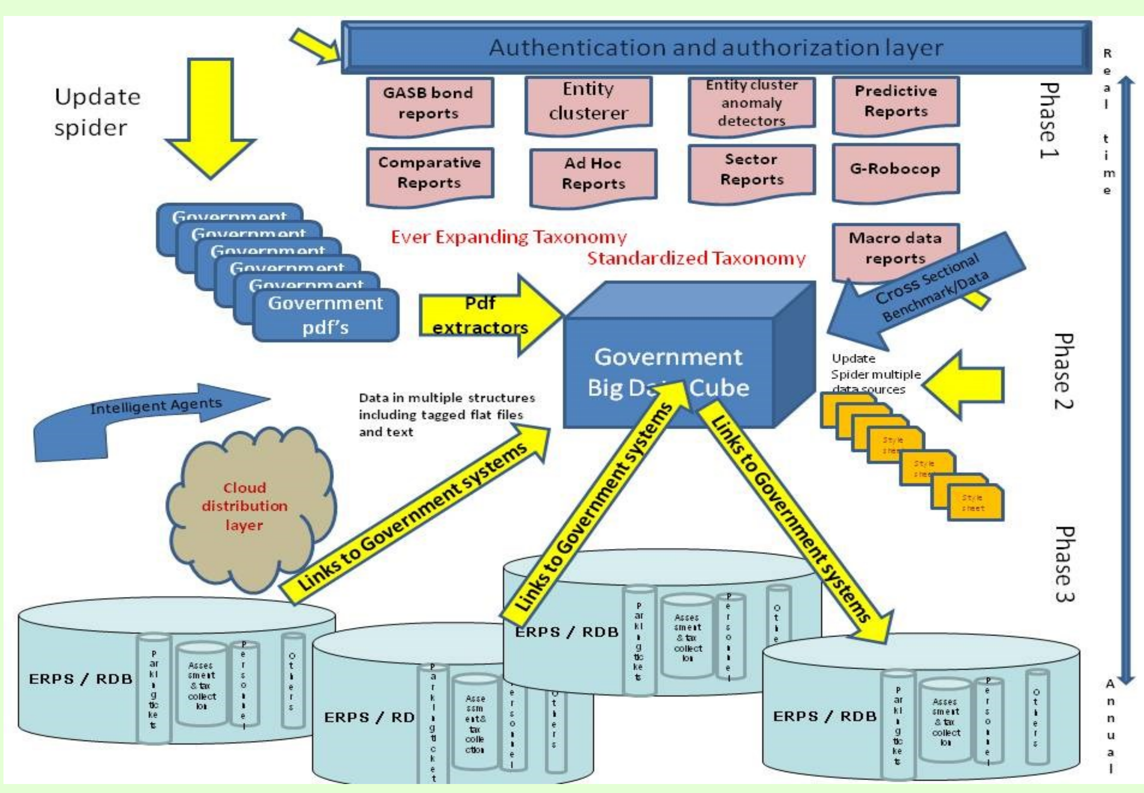
Rutgers Business School
Newark and New Brunswick

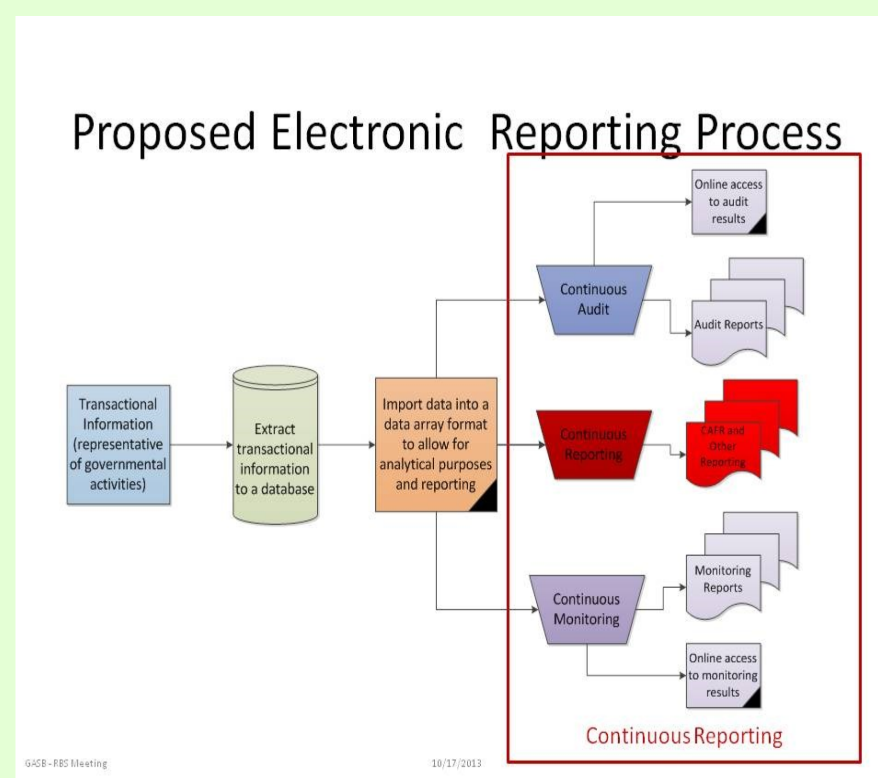
Selected CAR Lab Projects

GASB Project

Deniz Appelbaum, Desi Arisandi, Stephen Kozlowski, & Qiao Li

Advisors: **Irfan Bora, Hussein Issa, & Miklos Vasarhelyi**





Audit Data Standard– Procure to Pay

Tiffany Chiu, Jun Dai, & Joel Pinkus

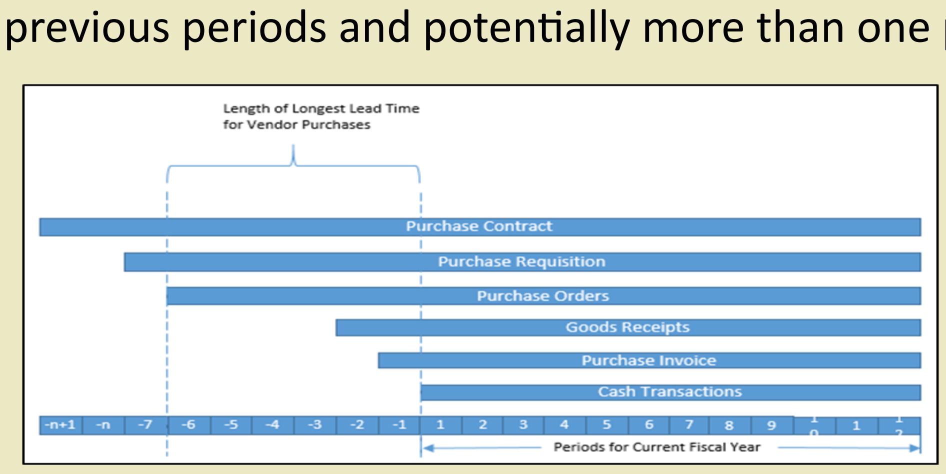
- The Procure to Pay (P2P) cycle effectively aligns the purchasing and accounts payable function of an organization.
- The process steps are illustrated below:

Purchase Contract → Purchase Requisition → Purchase Order → Goods Receipt → Price / Quantity Update or Returns → Purchase Invoice → Cash Disbursement

- The main focus of audit work conducted in the P2P cycle is centered on cash disbursements and normally includes control procedures for the 3-way match of Purchase Orders to Goods Receipts to Purchase Invoices.

Process Step	Definition
Purchase Contract	Agreement made with vendor for a purchase of product/services that will be consumed over a period of time.
Purchase Requisition	Internal Request for purchase of a product/services.
Purchase Order	Request of product/services from a vendor.
Goods Receipt	Notification of product receipt.
Price / Quantity Update or Returns	Adjustment of receipt based on prior agreement.
Purchase Invoice	Vendor invoice for product/services.
Cash Disbursement	Payment for product/services to vendor.

- The timeframe necessary to provide a complete picture of the P2P transactions for the current fiscal year may extend significantly into previous periods and potentially more than one previous period.



- Since the Purchase Contract and Purchase Requisition may not directly reflect the cash and 3-way match process, these process steps will be eliminated from the scope of the data extract, thus yielding the steps illustrated below:

Purchase Order → Goods Receipt → Price / Quantity Update or Returns → Purchase Invoice → Cash

Audit & Control Risk Assessment

Hussein Issa & Alexander Kogan

- Objective: Quality review of control risk assessment, learning tool for non-experts, risk-based sampling
- Methodology: Ordered Logistic Regression
- Software: SAS
- Model: Risk Level = f (Critical + Major + Non-Major)
- Dataset: Issues identified and overall risk assessment score

Identify Issues related to business processes → Classify Issues as Critical, Major, or Non-Major (CC, MC, NMC) → Assign an aggregate risk level for the overall process (L,M,H)

Ranking of Outliers

Ratio =

Calc.prob_Assigned Class
____
Calc. prob_Predicted Class

Difference = Calc. prob_Predicted Class—Calc.prob_Assigned Class

The lower the ratio, the more anomalous that record is

The larger the difference, the more anomalous the record is.

Process Mining

Michael Alles & Mieke Jens

Objective: Extract knowledge from event logs recorded by an information system and provide techniques and tools for discovering process, control, data, organizational, and social structures from event logs.

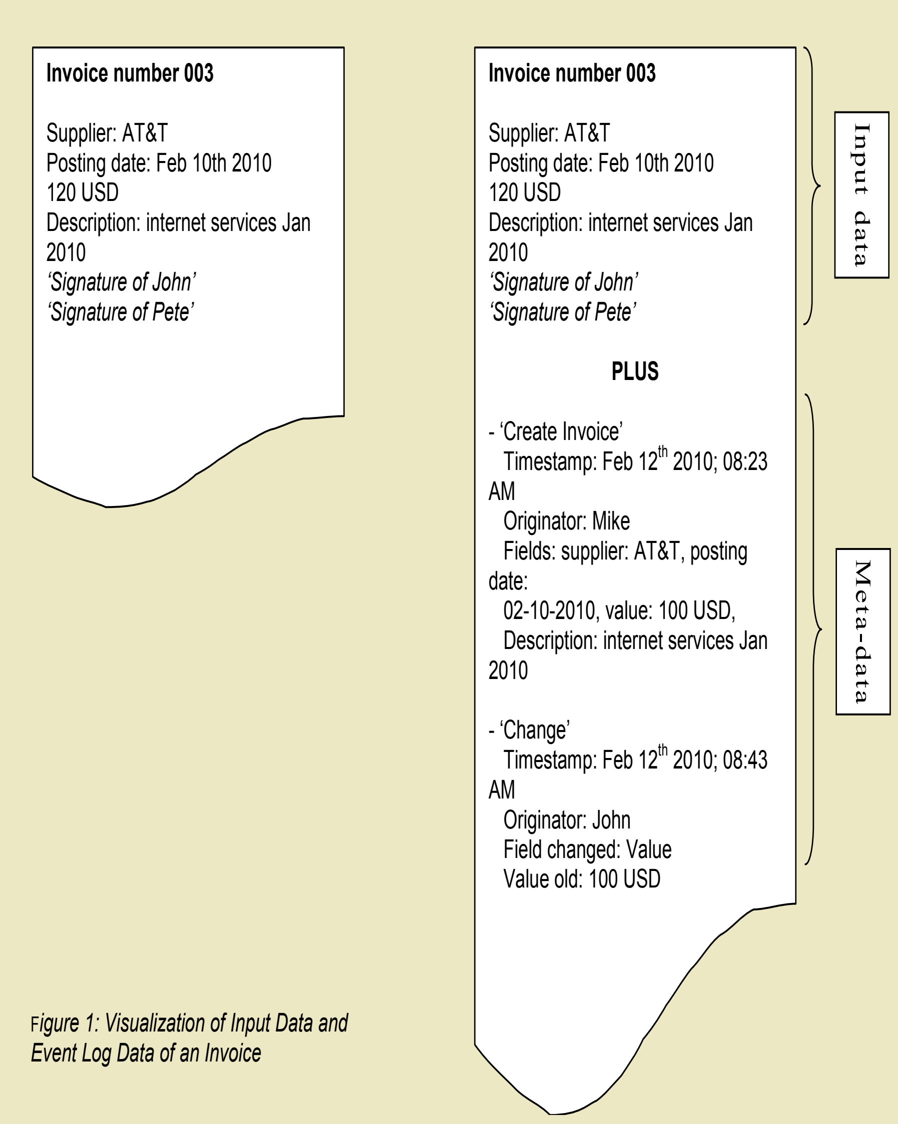


Figure 1: Visualization of Input Data and Event Log Data of an Invoice

The Development and Intellectual Structure of Continuous Auditing Research

Victoria Chiu, Qi Liu, and Miklos Vasarhelyi

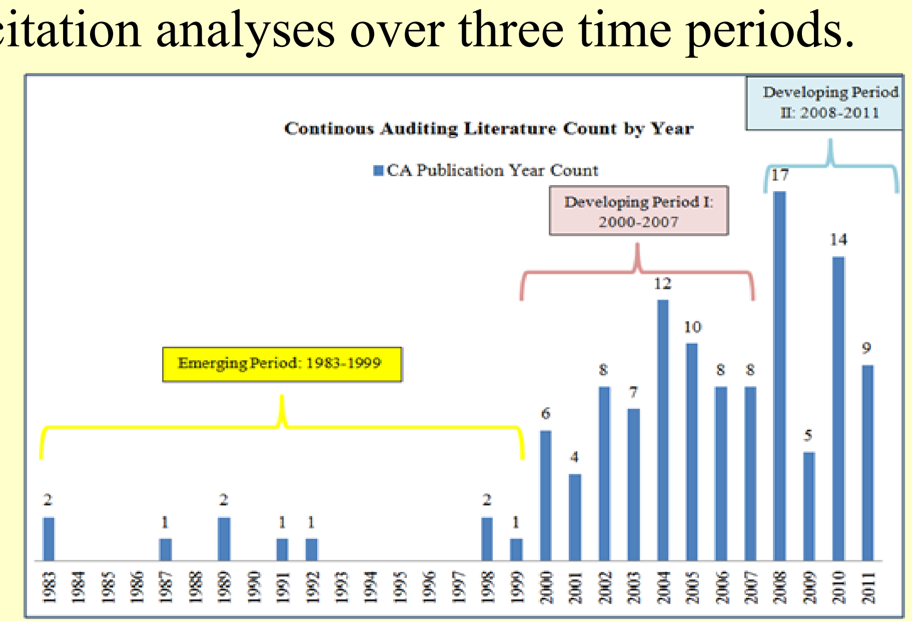
Introduction

The advancement in technology has been viewed as a significant force that influences the accounting profession (AICPA 1998). Auditing tasks has been evolving substantially by progressively utilizing the latest technology to improve process and procedure efficiency and effectiveness.

This study provides an examination on the development and citation's intellectual structure of continuous auditing research by (a) classifying continuous auditing research on the basis of four taxonomic categories, (b) applying citation and co-citation analyses to identify influential research and scholars within CA field, and (c) revealing main citations clusters that contribute to the formation of the continuous auditing field through the application of bibliometrics and graphical data mining techniques.

Methodology

A total of 118 continuous auditing research published from 1983 to 2011 were retrieved from online academic databases *(e.g., Ebsco-Host, Science Direct, Scopus...etc)* after querying key terms *(i.e. continuous auditing, continuous assurance, continuous monitoring, and continuous reporting).* A Continuous Auditing Taxonomy is developed to identify CA research characteristics — *Topical Area, Research Methods, Specific Area of Emphasis* and *Geographical Area* (Brown & Vasarhelyi 1994; Kogan et al. 1999). The intellectual structure of CA studies were analyzed by citation and co-citation analyses over three time periods.



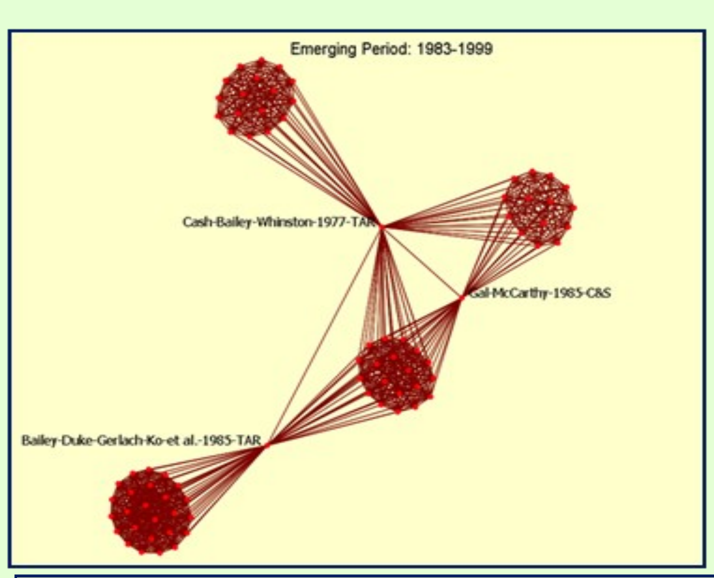
CA Research Characteristics

Topical area	Explanation
General CA	Involves high level CA research like grounding theory and framework of CA
Continuous Assurance	Includes studies about Continuous Data Assurance (CDA), (Continuous) System Assurance, Continuous Online Assurance and external Continuous Assurance research
Continuous Control Monitoring (CCM)	Consists research related to (continuously) monitoring of internal controls
Continuous Reporting (CR)	Denotes studies of frequent reporting and disclosure
Continuous Risk Monitoring and Assessment (CRMA)	Indicates dynamic risk measurement research
Enabling Technology	Refers to the essential technologies supporting CA and CR, such as electronization and XBRL
Audit Automation	Indicates studies on automating traditional manually performed auditing procedures.

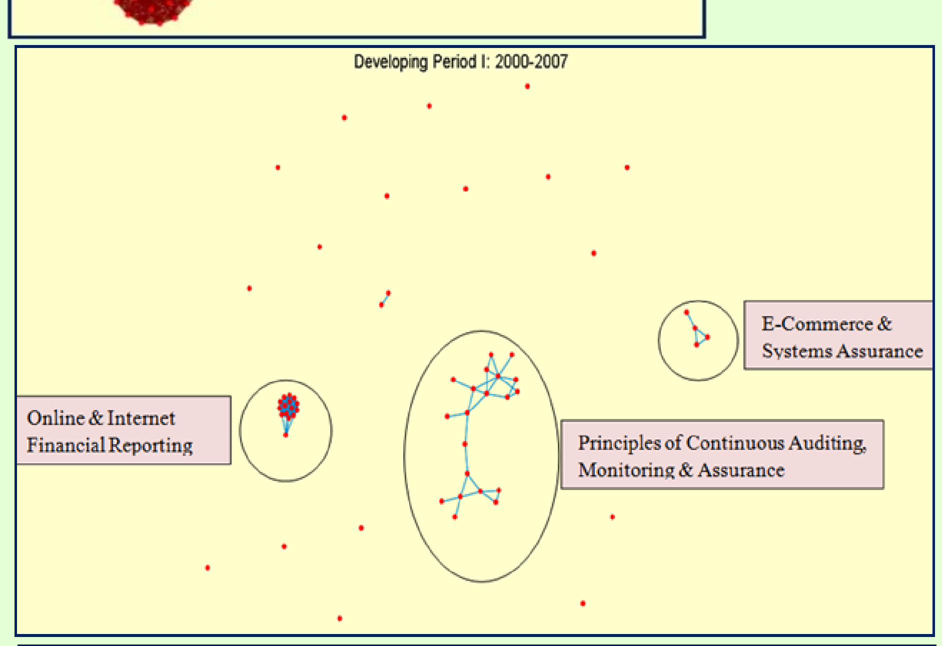
Topical Area /Research Method	Analytical	Archival	Experimental /Behavior	*Total*
General CA	31	1	8	*40*
Continuous Control Monitoring	6	2	12	*20*
Enabling technology	9	2	7	*18*
Continuous Reporting	3	5	6	*14*
Continuous Assurance	3	1	9	*13*
Audit Automation	6	0	5	*11*
Continuous risk monitoring and assessment	1	0	1	*2*
Total	*59*	*11*	*48*	*118*

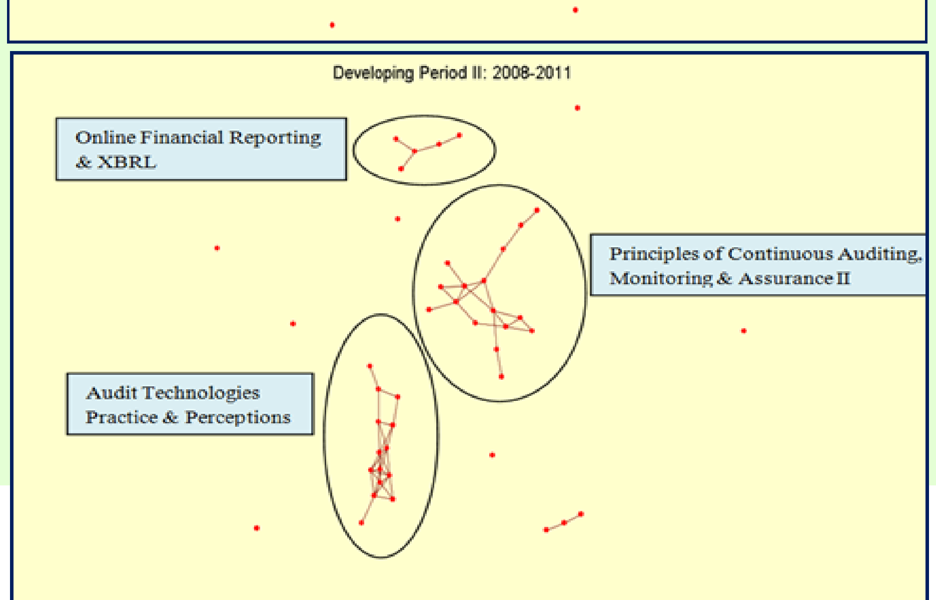
Topical Area\Specific Area of Emphasis	Architectural Issues Relating to CA	Effects/Consequences of CA	Factors Affecting CA	*Total*
General CA	17	17	6	*40*
Continuous Control Monitoring	12	7	1	*20*
Enabling technology	11	1	6	*18*
Continuous Reporting	3	4	7	*14*
Continuous Assurance	7	4	2	*13*
Audit Automation	6	1	4	*11*
Continuous risk monitoring and assessment	1	0	1	*2*
Total	*57*	*34*	*27*	*118*

CA Intellectual Structure



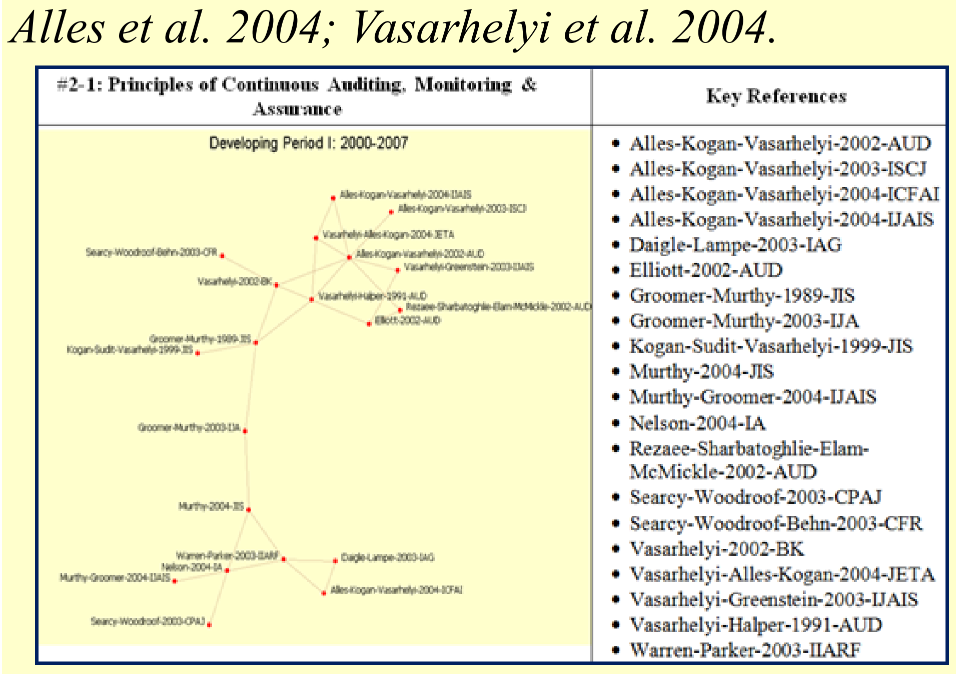
Citations Structure in
1983-1999
2000-2007
2008-2011

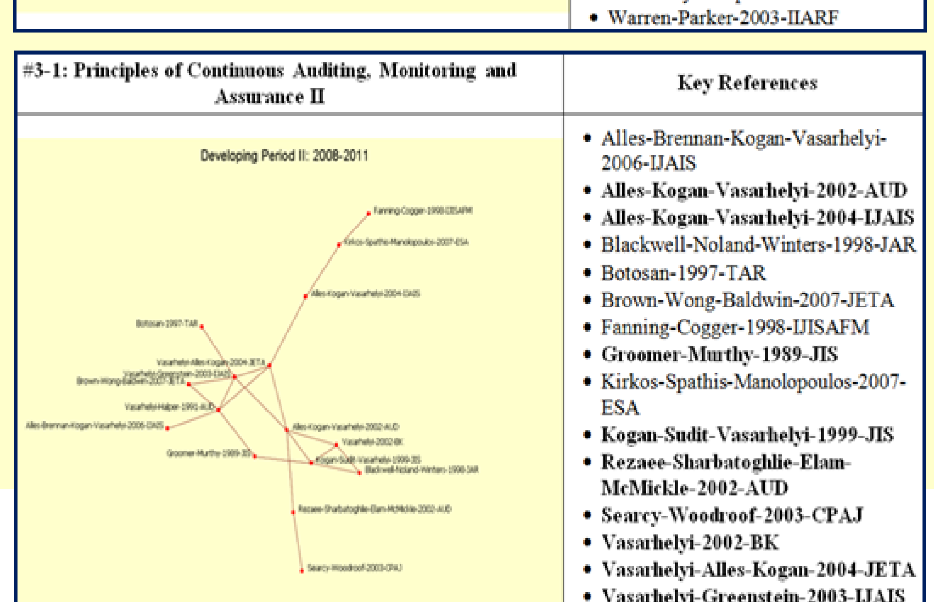




CA Intellectual Structure

Top ten influential manuscripts recognized in "*Principles of Continuous Auditing, Monitoring & Assurance"* clusters of period I & II — *Groomer & Murthy 1989; Vasarhelyi & Halper 1991; Kogan et al. 1999; Alles et al. 2002; Rezaee et al. 2002; Vasarhelyi 2002; Searcy & Woodroof 2003; Vasarhelyi & Greenstein 2003; Alles et al. 2004; Vasarhelyi et al. 2004.*





The Future of Continuous Audit

The three periods of research on CA evaluated in this paper show the absorption of technology into business, use of technology in auditing, and utilization of technology as an assurance tool.

Rapid change of technological enablement is driving the need for rapid knowledge development and leading to the obsolescence of traditional audit methods (Titera 2013). The original illustration of CA at Bell Labs (Vasarhelyi and Halper 1991) used primitive communication networks (RJE stations, print images, e-mail), limited computational power, and traditional assurance methods. Large data populations, computer-based processes, and a preponderance of automatic data collection are making manual auditing methods impossible. Research is needed to formalize accounting, analytic methods, and audit (Krahel 2012; Krahel and Vasarhelyi 2011). This automation must also be reflected on published accounting and audit standards (Titera 2013: Zhang et al. 2011). Much research is needed relative to (continuous) audit of Big Data, E-Commerce, transaction level XML, intelligent agents, textual analysis, etc.

RUTGERS
Rutgers Business School
Newark and New Brunswick