

KeAi

ISSN 2405-9188

*The Journal of*  
**FINANCE AND  
DATA SCIENCE**

Volume 1 • Number 1 • December 2015



Production and hosting  
by Elsevier on behalf of  
KeAi



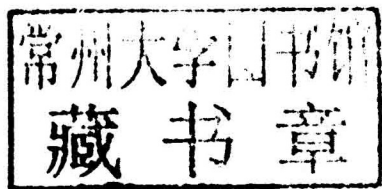
SCIENCE PRESS  
Beijing



Southwest Jiaotong University

# The Journal of Finance and Data Science

Volume 1, Number 1, December 2015



Production and Hosting by Elsevier on behalf of KeAi

# Aims and Scope

*The Journal of Finance and Data Science* (JFDS) will be the leading analytic journal on Finance and Data Science providing detailed analyses of theoretical (empirical) foundations and how they are used in financial economics in practice. JFDS gives the testing of well-established or new theories using financial data, the data-scientific measurement of variables relevant in financial decision-making and financial service, the econometrics of financial market data or the development of new econometric methodology with finance applications.

JFDS also gives the readers results-oriented computer hardware and software treatments and enhances full treatment of data information on practical finance products and techniques. Big Data Financial Economics Analytics raise new challenges and advanced state-of-the-art of computer science, and vice versa that computer data science provides emerging indispensable techniques in financial analysis. JFDS will include articles on

- Machine-learning, high frequency trading among algorithm trading in finance
- Theoretic and Empirical financial results from data science perspective
- Latest practical treatments on Data Science in financial economics
- Innovative designs and technique tools for computer hardware and software for finance
- How to apply theoretic results into real world problems related to Data Science
- Illustration and rigorous analysis of essential innovations in Data Science on Financial Economics
- New financial products modeled by data science
- Case study on industry to the financial services and financial economics related to Data Science and traditional financial research areas.

# The Journal of Finance and Data Science

## Editorial Board

### Editor-in-Chief

**Weiping Li**

*Department of Finance,  
The Spears School of Business,  
Oklahoma State University, US*

### Editorial Board

**Stephen Blyth**

*Harvard University, US*

**Damiano Brigo**

*Imperial College London, UK*

**Andreas Buja**

*University of Pennsylvania, US*

**Julian R. Franks**

*London Business School and  
Academic Director of the Centre for  
Corporate Governance, UK*

**Lars Peter Hansen**

*University of Chicago, US*

**Dengshi Huang**

*Southwest Jiaotong University, China*

**Jingzhi (Jay) Huang**

*Penn State University, US*

**Ohad Kadan**

*Washington University in St. Louis, US*

**Petter Kolm**

*New York University and the Principal of the  
Heimdall Group, LLC. US*

**Ed Lazowska**

*University of Washington, US*

**Yuying Li**

*University of Waterloo, Canada*

**Andrew W. Lo**

*Massachusetts Institute of Technology, US*

**Massimo Massa**

*INSEAD, France*

**Lubos Pastor**

*University of Chicago, US*

**Hanspeter Pfister**

*Harvard University, US*

**Vladimir Piterbarg**

*Barclays Capital, UK*

**Artur Sepp**

*Bank of America  
Merrill Lynch, UK*

**Neil Shephard**

*Harvard University, US*

**Robert J. Shiller**

*Yale University, US*

**Jian Wang**

*Alibaba, China*

**Yanwu Yang**

*Huazhong University of Science and  
Technology, China*

**Kathy Yuan**

*London School of Economics and  
Political Science, UK*

**Jianfeng Zhan**

*Chinese Academy of Sciences, China*

**Aoying Zhou**

*East China normal University, China*

# The Journal of Finance and Data Science

Volume 1, Number 1, December 2015

## Table of Contents

### Original Articles

Big data based fraud risk management at Alibaba <i>Jidong Chen, Ye Tao, Haoran Wang, Tao Chen</i>	1
A case study on loan loss analysis of a community bank <i>Weiping Li, Philip McMahan</i>	11
Latency critical big data computing in finance <i>Xinhui Tian, Rui Han, Lei Wang, Gang Lu, Jianfeng Zhan</i>	33
Required Market Risk Premium among countries in 2012 <i>Pablo Fernandez, Javier Aguirreamalloa, Isabel Fernandez Acín</i>	42



# Big data based fraud risk management at Alibaba

Jidong Chen\*, Ye Tao, Haoran Wang, Tao Chen

*Alipay, Alibaba, China*

Received 15 February 2015; revised 8 March 2015; accepted 20 March 2015

Available online 14 July 2015

## Abstract

With development of mobile internet and finance, fraud risk comes in all shapes and sizes. This paper is to introduce the Fraud Risk Management at Alibaba under big data. Alibaba has built a fraud risk monitoring and management system based on real-time big data processing and intelligent risk models. It captures fraud signals directly from huge amount data of user behaviors and network, analyzes them in real-time using machine learning, and accurately predicts the bad users and transactions. To extend the fraud risk prevention ability to external customers, Alibaba also built up a big data based fraud prevention product called AntBuckler. AntBuckler aims to identify and prevent all flavors of malicious behaviors with flexibility and intelligence for online merchants and banks. By combining large amount data of Alibaba and customers', AntBuckler uses the RAIN score engine to quantify risk levels of users or transactions for fraud prevention. It also has a user-friendly visualization UI with risk scores, top reasons and fraud connections.

© 2015, China Science Publishing & Media Ltd. Production and hosting by Elsevier on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Fraud detection and prevention; Risk model; Malicious behavior; Risk score; Big data analysis

## 1. Introduction

Big data is an all-encompassing term for any collection of data sets which are large, complex and unstructured, so that it becomes difficult to process using traditional data processing applications. Big data “size” is dynamic and constantly growing, as of 2012 ranging from a few dozen terabytes to many petabytes of data by the time when the article is written. It is also a set of techniques and technologies that to analyze, capture, curate, manage and process data within a tolerable elapsed time (Wikipedia).

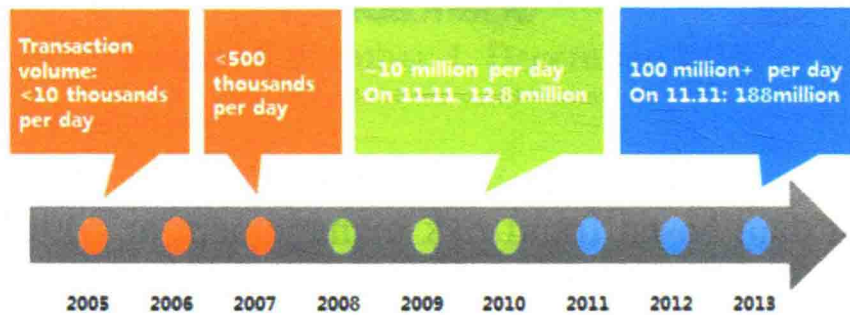
Big data has many different purposes – fraud risk management, web display advertising, call center optimization, social media analysis, intelligent traffic management and among other things. Most of these analytical solutions were not possible previously because data technology were unable to store such huge size of data or processing technologies were not capable of handling large volume of workload or it was too costly to implement the solution in a timely manner.

\* Corresponding author.

E-mail addresses: [jidong.cjd@alibaba-inc.com](mailto:jidong.cjd@alibaba-inc.com), [chenjd@gmail.com](mailto:chenjd@gmail.com) (J. Chen).

Peer review under responsibility of China Science Publishing & Media Ltd.





Graph 1. Transaction volume increases at Alibaba.

With business needs arose, Alibaba employed optimized system and platform and developed advanced methodology and approach to handle 10 billion level daily volume. It started from data platform of RAC in 2009, via GP (Green Plum, an EMC product, see EMC<sup>2</sup>) and Hadoop (see White<sup>7</sup>), and is using ODPS now. Data processing and analyzing is also improved from  $T + 1$  mode<sup>1</sup> to near real-time mode.

By adapting big data techniques, Alibaba highlights advances made in the area of fraud risk management. It invents a real-time payment fraud prevention monitoring system, called CTU (Counter Terrorist Unit). And CTU becomes one of the most advanced online payment fraud management system in China, which can track and analyze accounts' or users' behavior, identify suspicious activities and can apply different levels of treatments based on intelligent arbitration.

Fraud risk models are one of the supportive layers of CTU<sup>2</sup> (Counter Terrorist Centre). They use statistical and engineering techniques to analyze the aggregated risk of an intermediary (an account, a user or a device etc). Detailed attributes are generated as inputs. Different algorithms are to assess the correlations of these attributes and fraud activities, and to separate good ones from bad ones. Validating and tuning are to make sure models apply to different scenarios. Big data at Alibaba produces thousands and thousands of attributes, and fraud risk models are built to deal with various of fraud activities.

Those big data based fraud models are widely used in almost every procedure within Alibaba to monitor fraud, such as account opening, identity verification, order placement, before and after transaction, withdrawal of money, etc. To build up a safe and clean payment environment, Alibaba decides to expand this ability to external users. A user-friendly product is built, called AntBuckler. AntBuckler is a product to help merchants and banks to identify cyber-crime risks and fraud activities. And a risk score (RAIN Score) is generated based on big data analysis and given to merchants and banks to tell the risk level.

In this paper, we show that Alibaba applies the big data techniques and utilizes those techniques in fraud risk management models and systems. We also present the methodology and application of the big data based fraud prevention product AntBuckler used by Alibaba.

The remainder of the paper is organized as follows. Section 2 introduces big data applications and basic computing process at Alibaba. Section 3 explains fraud risk management at Alibaba in details and fraud risk modeling. Section 4 provides an explanation of AntBuckler. We conclude in Section 5.

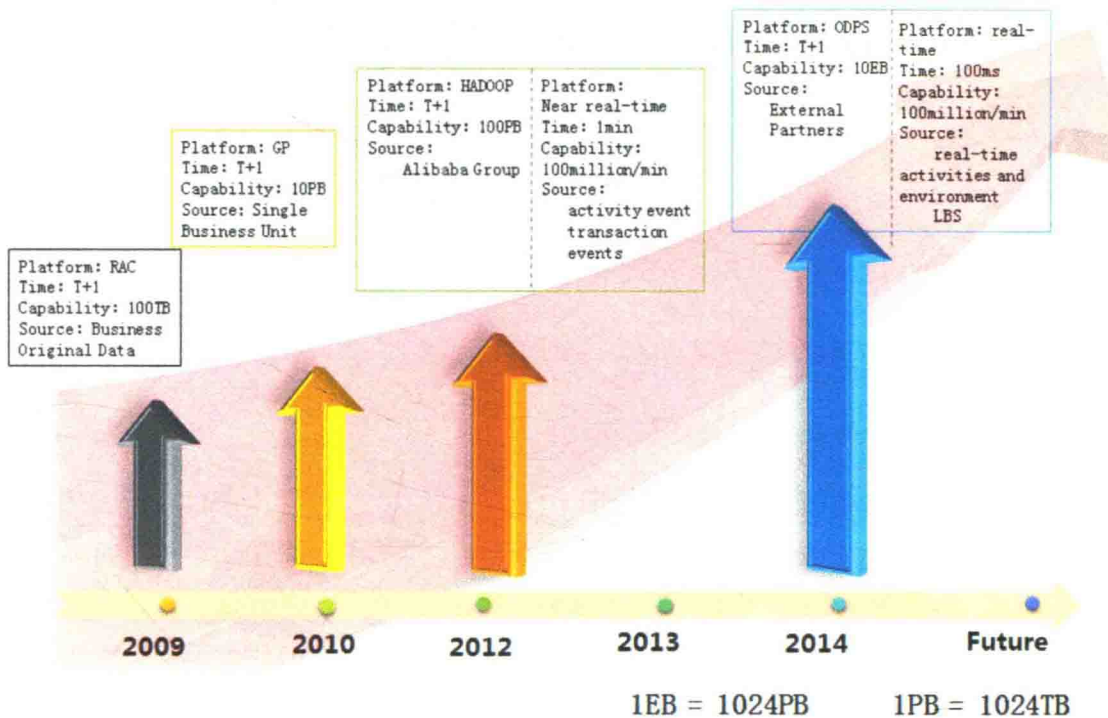
## 2. Big data applications at Alibaba

Alibaba grows fast in past 10 years. In 2005, daily transaction volume was less than 10 thousands per day. It reached to 188 million on Nov 11th, 2013 one day. Graph 1 illustrates that the transaction volume at Alibaba changes from 2005 to 2013 on daily basis.

With business growing exponentially, data computing, processing system and data storage are bound to change as well. It started from data computing platform of RAC (Oracle Real Application Clusters (see Oracle white paper<sup>1</sup>)) in 2009, via GP and Hadoop, and is using ODPS now. Data processing and analyzing is also improved from  $T + 1$  mode

<sup>1</sup>  $T + N$  mode:  $T$  is time, when system runs.  $N$  is time interval.  $T + 1$  means the system runs on the second day.

<sup>2</sup> CTU: Alipay's internal risk control system, which is fully developed and designed by Alibaba. This name is inspired by American TV series <<24>>.



Graph 2. Big data computing progress at Alibaba.

to real-time mode, especially in risk prevention at Alibaba, fraud check on each transaction can be controlled within 100 ms (millisecond). Moreover, data sources are extended from single unit data to a combination of internal group data and external bureau data. Graph 2 illustrates that the big data computing process at Alibaba progress extensively since 2009. Alibaba has data not only from Taobao, Tmall and Alipay, but also from partners like Gaode Maps and others. Data from various sources builds up an integrated data platform, the platform from business scenarios extends largely as well. Marketing uses data analysis to target users accurately and to provide customers service personally. Merchants and financial companies need professional data classification to filter out valuable customers. Intelligent customer service can effectively and efficiently solve users' requests and complaints using comprehensive data platform. And the online payment services and systems, Alibaba, among leaders of online payment service providers, builds up a fraud risk management platform to ensure both buyers and sellers with fast and safe transactions. Alibaba deals with big data extensively on credit score and insurance prices as well as other types of business.

### 3. Fraud risk management at Alibaba

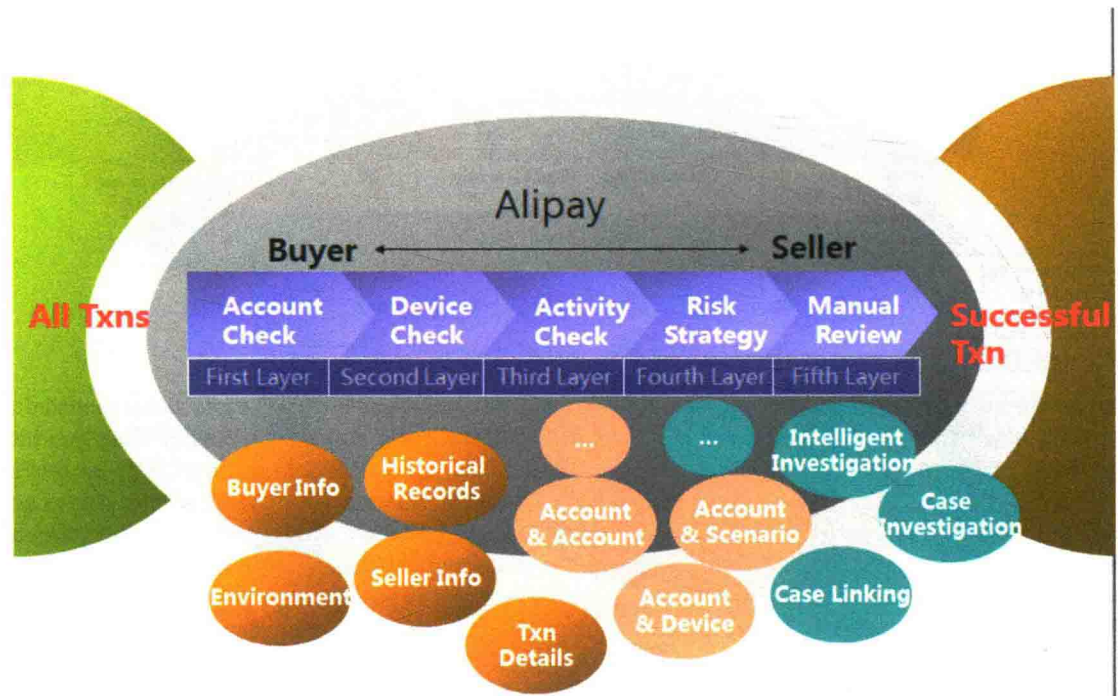
#### 3.1. Framework for fraud risks

Fraud risk management at Alibaba is totally different from that of traditional financial and banking system now due to big data. To deal with real-time frauds, new engineering approaches are gradually developed to handle such quantity of data. On top of hardware system, risk prevention framework is also built up to support new methodology and algorithm. There are a few different kinds of risk prevention frameworks.

One fundamental framework of fraud risk that Alibaba uses is called multi-layer risk prevention framework. Graph 3 illustrates the multi-layer risk prevention framework Alibaba uses in Alipay System. There are total five layers in this system.

At Alibaba, there are 5 layers to prevent fraud for a transaction. The five layers are (1) Account Check, (2) Device Check, (3) Activity Check, (4) Risk Strategy and (5) Manual Review. One fraudster can pass first layer on account check, and then there are still four layers ahead to block the fraudster. When a transaction is initiated, the first layer is





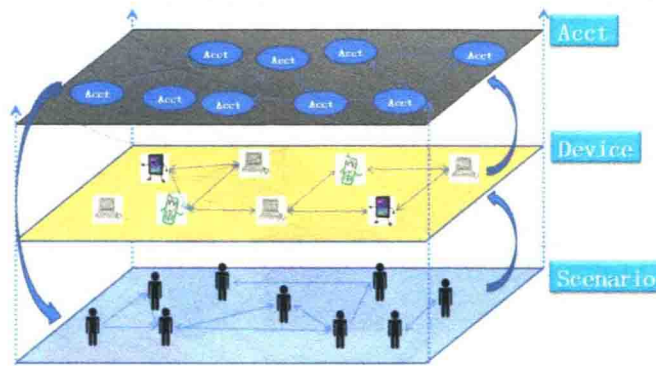
Graph 3. Multi-layer risk prevention framework.

Account Check, which includes buyer account information and seller account information. Several checks on the first layer Account Check are designed as questions: does the buyer or seller account have bad/suspicious activity before? Is there any possibility the buyer account stolen etc? Extremely suspicious transactions may be declined to protect genuine buyers, or extra authentic methods may be triggered to double confirmation in this situation. The second layer is Device Check, which includes the IP address check and operation check on the same device. Similarly, checks on the second layer Device Check are designed from passing several questions: whether there are huge quantify of transactions from the same device? Any transaction is from bad devices? The third layer is Activity Check, called as Behavior Check as well, which checks historical records, buyer and seller behavior pattern, linking among accounts, devices and scenarios. Checks on the third layer Activity Check are designed as questions as well: whether the buyer or seller account link to an identified bad account? The fourth layer is Risk Strategy, which makes final judgment and takes appropriate action. Checks on the fourth layer Risk Strategy are designed to aggregate all results from previous checks according to severity levels. Some transactions are sent to auto-decision due to obvious fraud activities. Some grey cases are sent to Manual Review. On one hand, Alipay would like to provide better services and experiences for both parties. On the other hand, Alipay does not want to misjudge any case. Without strong evidence, suspicious cases will be manually reviewed in the last layer Manual Review, where more evidences are revealed and some phone calls may be made to verify or remind or check with buyers or sellers.

Another key difference between fraud risk management at Alibaba and “that” of traditional financial and banking system is the risky party. Customers are evaluated to be the main risky party in banking system. At Alibaba, there are 3 layers of risky party. The three layers are (1) Customer level, (2) Account level and (3) Scenario level. See Graph 4. Risk fraud prevention at Alibaba is for both customers whether they are buyers or sellers or not, for both accounts whether those accounts are prestigious for big company or a single individual or not, for both scenarios whether those activities happen during account opening or money withdrawal.

### 3.2. CTU – fraud prevention monitoring system

CTU is a real-time payment fraud prevention monitoring system, which can track and analyze accounts' or users' behavior, identify suspicious activities and apply different level of treatments based on intelligent arbitration. The first



Graph 4. Multi-layer risk prevention framework.

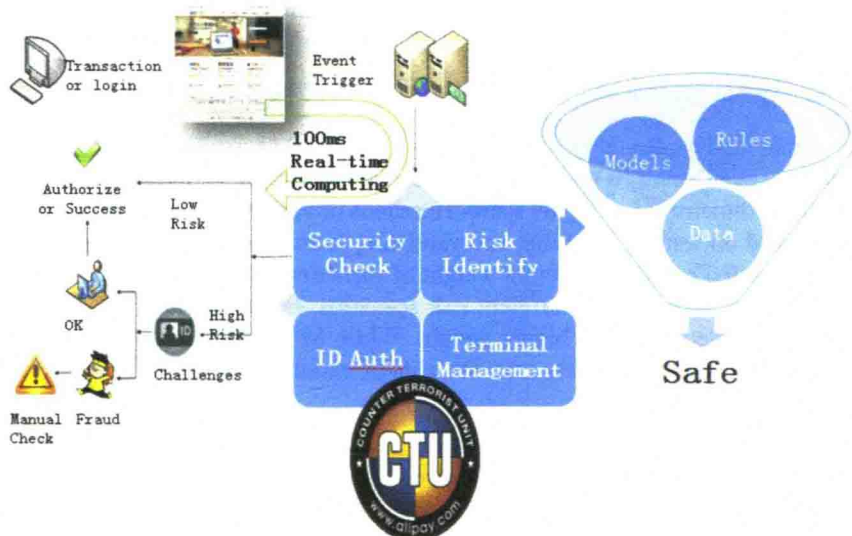
version launched on 1st Aug, 2005. This system is independently developed by Alipay's risk control team. At that time, it focused more on large transaction investigation, suspicious refund, etc. Now it extends to money laundry, marketing fraud, accounts, and card stolen/loss as well as cash monetization. Additionally, it is a 24 h monitoring system, which provides throughout protection at any time.

When an event happens, it passes through CTU for judgment. An event is defined to be that a user logs in, changes profiles, initiates transactions, withdraws money from Alibaba to other bank accounts and others. There are hundreds of kinds of events. A suspicious event triggers models and rules behind the CTU for real-time computing, and within 100 ms, CTU returns the result with risk decision. If this is a low risk from CTU return, the event is passed to continue its operation. If this is a high risk, the CTU will direct a stop or a further challenge step to continue the process. Graph 5 illustrates the CTU operating process.

### 3.3. Fraud risk modeling

The data that supports CTU judgment is from historical cases, user behaviors, linking relationship and so on. Risk models are built to analyze fraudsters' cheating patterns, relations among fraudsters, different behaviors between a group of good users and a group of bad ones.

There are a few factors to consider during building risk models. Bias and Variance are usually concerned together to balance the effectiveness and impact of a risk model. Bias is a factor to measure how a model fits for the risk, how to



Graph 5. CTU-risk prevention system at Alibaba.





Graph 6. Three dimension of RAIN score.

find the risk of account or transaction accurately. Variance is to measure whether a model is stable or not, whether it can sustain a relative longer lifecycle in business. Negative positive rate, also called wrong cover rate, is to measure how the accuracy a model is. High negative positive rate will bring company huge business pressure and bad user experience. Moreover, interpretability is necessary to explain to users the reason that a model gives such risk level to his account or transaction. In big data age, except for factors mentioned above, data scientists keep fighting for data deficiency, data sparsity and data skewness.

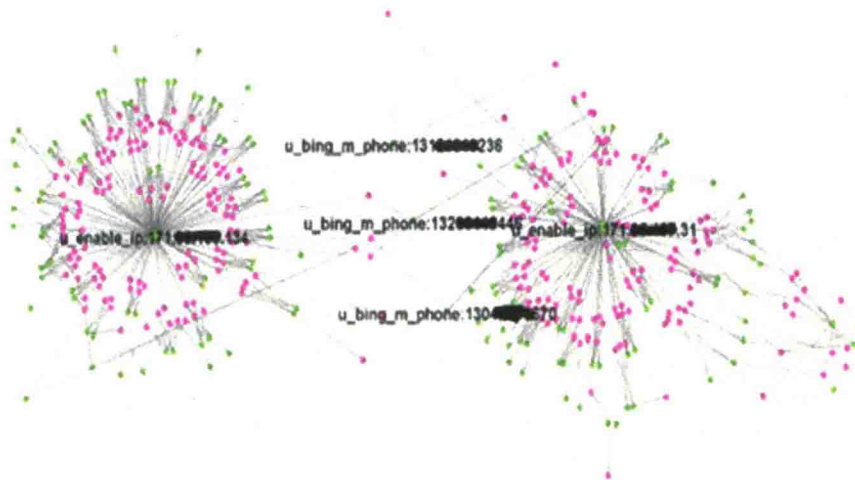
Model building process is also relatively mature at Alibaba after repeated deliberation. White and black samples are chosen first. White samples are good risk parties. Black samples are usually risky parties judged as bad. A good model can differentiate white and black samples to the most degree. Behavior data and activity data are collected for both samples to generate original variables from abstracting aggregated variables. Through testing, some variables are validated effectively. They can be finally used in model building. From our modeling experience by using Alibaba big data, decision tree C5.0 and Random Forest have better performance to balance between Bias and Variance. One obvious reason is that they do not assume data distribution since they are algorithmic models rather than data models. When a model is able to better separate good and bad ones among samples, the model is basically adapted to process. However, to make sure it applicable to different scenario, validation is also important. A model can be launched if it works effectively and efficiently on testing and validating data. Then, the fraud risk prevention models are needed to be deployed in the production environment, and are used in CTU combining with other strategies and rules.

### 3.4. RAIN score, a risk model

RAIN is one kind of risk models. RAIN stands for Risk of Activity, Identity and Network. Basically, the risk of an object (a user, an account or even a card) is composited of three dimensions of variables, activity, identity and network. Graph 6 illustrates the three dimensions of RAIN score. Hundreds of variables are first selected to interpret status and behavior of an object. Based on testing, verifying and validating from fraud risk models, variables are selected and kept. A RAIN score is generated based on different weight of variables within these three dimensions. Variables and weight of variables may differ according to different risk scenarios. For example, for a card stolen scenario, more Identity variables may be selected and with higher weight. While for a credit speculation scenario, more Network variables may be selected and are given higher rate. Weights of variables are trained by different machine learning algorithms, such as logistic regression.

### 3.5. An example of network-based analysis in fraud risk detection

Graph theory (Network-based Analysis), an applied mathematical subject is commonly applied on social network analysis (see Wasserman<sup>6</sup>). Facebook, Twitter apply the Graph theory on their social network analysis. Network-based



Graph 7. Network of accounts and information.

analysis plays a new role in the risk control. Fraudsters nowadays are mischievous. They know that online risk models are constantly checking whether fraud accounts are from same name, address, phone and credit card, etc. Hence, they try new ways to hide the connections. Therefore, network based analysis is introduced to reveal the connection in this area. For example, if each account is considered as a node, network-based analysis is to locate edges among different nodes if they are owned by a physical person. If there are reasonable ways to define the edges among different nodes, some interesting groups can be disclosed.

In Graph 7, the red nodes represent accounts and green nodes are detailed profile information for those accounts, such as enabled IPs, phone numbers, name, address, etc. If an account (red node) has detailed profile information (green node), network analysis draws a line between this red and green node to show the relationship and the line is an edge. Graph 7 illustrates the network analysis that both groups have their own enabled IPs. However some accounts in two groups shared same bind phone numbers. This exposes the connections between two groups. Another example as below.



Graph 8. Betweenness detection.



Graph 8 tells us another story. One account shares the same register IP and register device footprint with the left group. It also shares the same name and info number with the right group. This is a strong evidence to prove the connections between two groups of accounts.

Two examples above are just a simple case. In the real world, connections are extremely complicated. We have to use paralleled graph algorithms and special graph storage to handle the huge network connection graph. The betweenness nodes (concepts from network-based analysis, see Freeman<sup>5</sup>) play important roles in finding connections of different accounts, where betweenness nodes are the betweenness centralities used in the network analysis. Connections now is widely used to judge relationship of accounts, this effectively prevent fraudsters building up their own networks.

#### 4. AntBuckler – a big data based fraud prevention product

To build up a safe and clean payment environment, Alibaba decides to expand its risk prevention ability to external users. A big data based fraud management product is built and called AntBuckler. This product is fully developed by Alipay.

AntBuckler is a product to help merchants and banks to identify cyber-crime risks and fraud activities. We find that merchants generally deal with similar fraud patterns. One example is marketing program fraud. Merchants often give cash reward or voucher certificate to new users to expand their user base. Fraudsters often take this opportunity to create hundreds of different accounts. To merchants, the marketing resource is not given to correct user base. To good users, they cannot benefit the cash reward or voucher certificate. Fraudsters may also sell their accounts with coupon in higher price. This not only damages the merchant's brand image and reputation, but also confuses the market and potential customers.

Antbuckler uses the RAIN model engine and generates a risk score (RAIN Score) to quantify the risk level. The score ranges from 0 to 100. The higher, the riskier. It also has user-friendly visualization. Top reasons are shown on top with a higher weight and brighter colors. Connection, through account, email, phone, card and so on, are presented using network based view. See Fig. 1. Fig. 1 is the main interface of one risky account. The interface gives detailed

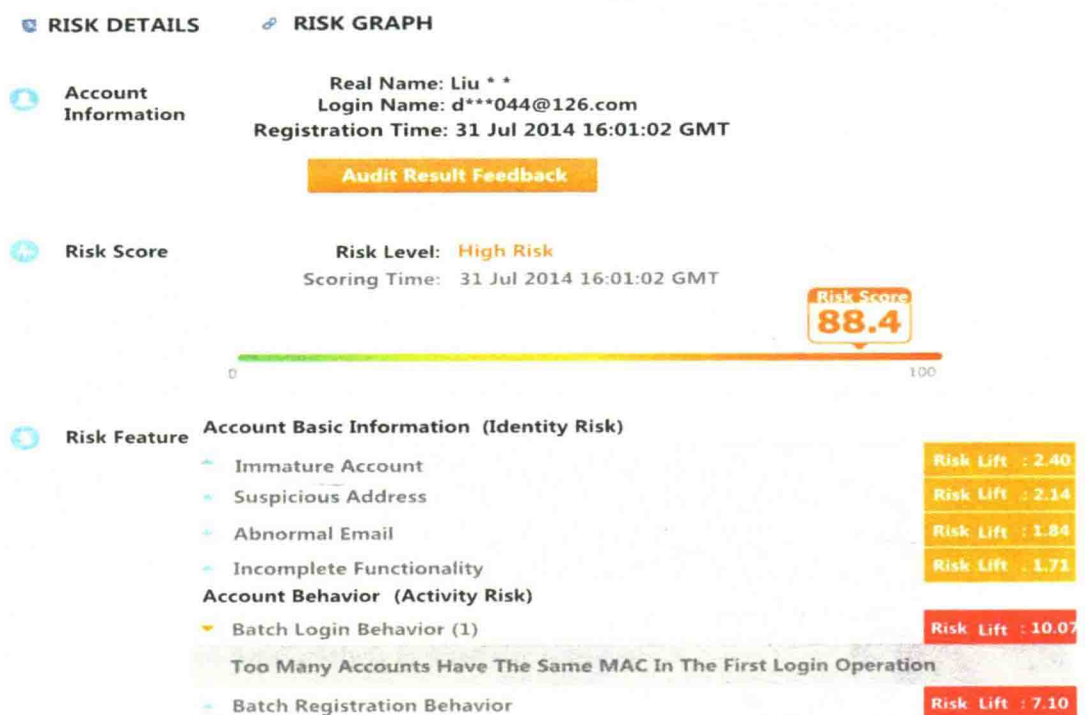
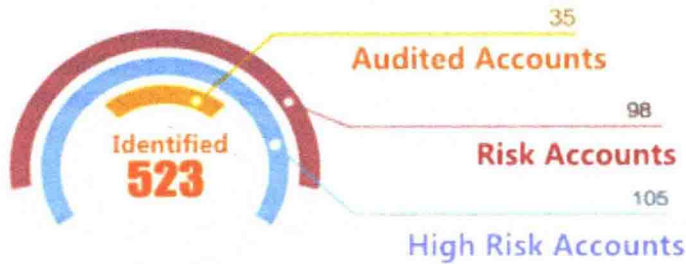


Fig. 1. User-friendly visualization of AntBuckler.

Select the Query Date: **TODAY** LATEST WEEK LATEST MONTH LATEST 3 MONTHS

Number of Risk Recognized **28888**



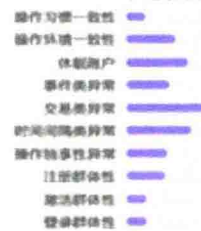
**Account Basic Information**  
Identity Risk



**Account Relationship**  
Network Risk



**Account Behavior**  
Activity Risk



账户名	风险评分	风险等级	评分时间	操作
da_cdfaca@163.com	97	最高	2014-02-02 19:20	<a href="#">查看详情</a>
2424432432@163.com	87	最高	2014-02-02 19:20	<a href="#">查看详情</a>
esarewa@126.com	82	最高	2014-02-02 19:20	<a href="#">查看详情</a>
13455555@163.com	68	高	2014-02-02 19:20	<a href="#">查看详情</a>

Fig. 2. Operation dashboard of AntBuckler.

account information, such as name, login email and registered time. RAIN score is shown together with a colorful bar. Green means safer and red means riskier. The operation dashboard tells how many accounts are judged by AntBuckler, how many accounts are identified with risk, where risky accounts are distributed, etc. See Fig. 2. Fig. 2 is an example of operation dashboard.

## 5. Conclusion

Big data based fraud risk management is a new trend in payment service business. It leads to a new generation of fraud monitoring and fraud risk management, which is based on big data processing and computing technology, real-time fraud prevention system and risk models. Alibaba has successfully used big data based fraud risk management to deal with daily fraud events. In this paper, we outline the fraud risk management at Alibaba and a big data based fraud prevention product. We would like to extend the analysis to build a safer and cleaner payment environment.

## References

1. An Oracle White Paper. *Oracle Real Application Clusters (RAC)*; 2013. <http://www.oracle.com/technetwork/products/clustering/rac-wp-12c-1896129.pdf>.
2. EMC Inc. *Greenplum Database: Critical Mass Innovation, Architecture White Paper*; 2010. <https://www.emc.com/collateral/hardware/white-papers/h8072-greenplum-database-wp.pdf>.
5. Freeman LC. Centrality in social networks I: conceptual clarification. *Soc Netw.* 1979;1:215–239.
6. Wasserman Stanley, Faust Katherine. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press; 1994.
7. White Tom. *Hadoop: the Definitive Guide*. O'Reilly Press; 2012.

# A case study on loan loss analysis of a community bank

Weiping Li <sup>a,b,\*</sup>, Philip McMahan <sup>c</sup>

<sup>a</sup> Southwest Jiaotong University, Chengdu, Sichuan Province, 610031, PR China

<sup>b</sup> Oklahoma State University, Stillwater, OK, 74078-0613, USA

<sup>c</sup> 5624 NW 130 St., Oklahoma City, OK, 73142, USA

Received 11 March 2015; revised 29 March 2015; accepted 30 March 2015

Available online 17 July 2015

## Abstract

In this paper, we perform a case study on a community bank from the aspect of the terms structure of the banks historical loan losses and the loan quality rating matrix used by the bank. The data source is from a small rural community bank located in a Midwestern state of the United States. The basic statistical analysis of the loan losses that is based on the banks internal loan quality rating system is presented.

© 2015, China Science Publishing & Media Ltd. Production and hosting by Elsevier on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*JEL classification:* G21; G24; G28; G33; G38

*Keywords:* Community bank; Loan grade; Facility rating matrix; Term structure of the loss

## 1. Introduction

The continued improvement of the credit analysis process of a bank is indispensable in the current financial environment for any bank, especially a community bank. Advances in information technology, the development of new financial instruments, and other innovations in banking, along with increased regulation and competition from non-regulated financial entities have made a much less friendly environment for community banks today than in the past.

During the most recent financial crisis, a large number of commercial banks in the United States were closed. The regulators closed 507 banks from 2008 to 2012, while only 25 banks were closed in the period 2001 through 2007. Commercial banks having less than one billion dollars in assets made up 85% of the number of banks closed between 2008 and 2014 and made up 92% of the number of banks closed from 2001 through 2007. The median level of leverage capital for the community banking industry as of June 2010 was 9.45%. This was roughly equal to the median

\* Corresponding author. Southwest Jiaotong University, Chengdu, Sichuan Province, 610031, PR China.

E-mail addresses: [w.li@swjtu.edu.cn](mailto:w.li@swjtu.edu.cn), [w.li@okstate.edu](mailto:w.li@okstate.edu) (W. Li), [golf73521@gmail.com](mailto:golf73521@gmail.com) (P. McMahan).

Peer review under responsibility of China Science Publishing & Media Ltd.



capital ratios of the banks that failed from 2008 thru 2009, 10 quarters before they failed. In those years, 165 banks failed with 123 (75%) having assets less than \$1 billion.<sup>d</sup>

Given these facts regarding the number of smaller bank failures as a percentage of the total number of bank failures, improvement in both the methodology of developing an adequate loan loss reserve and the credit risk management of these banks is necessary for the survival of smaller banks in the future.

This paper examines how a community bank might use a loan quality rating system which contains both objective factors and judgmental adjustments to first stratify the loans held by the bank into risk categories and then use these risk categories to develop a robust methodology to calculate a loan loss reserve for the bank. In addition, this data can be used to develop stress test models regarding loan losses for the bank. Stress testing for community banks was called for in a paper presented to the Conference of State Bank Supervisors in 2010. The term structure of the loan loss analysis for the loan quality ratings in our study are the basic mean and variance for each rating is given through the years 2007–2014 as the bank had data for that period.

This paper will focus only on what a community bank could do to use data already available to improve its credit risk management process. In order to focus on a community bank, it is important to understand what we define as a community bank. A community bank can be referred to as a bank having, typically, less than or equal to one billion dollars in total assets. On the other hand, this single proxy used to define a community bank does not take into account other factors which may also define a community bank. We would like to adapt a more complete definition of a community bank has given by DeYoung et al (2004)<sup>e</sup> as:

*A community bank holds a commercial bank or thrift charter; operates physical offices within a limited geographic area; offers a variety of loans and checkable insured deposit accounts; and has a local focus that precludes its equity shares from trading in well-developed capital markets.*

FDIC (2012)<sup>f</sup> refers to the problem of defining community banks by size due to some factors such as inflation, economic growth and the size of the banking industry varying over time. We perform this case study for a community bank with total assets less than \$1 billion.

The financial crisis of 2007–2009 lead to the widespread failure of financial institutions and/or the freezing of capital markets. The financial regulations and insolvency procedures for large complex financial institutions, including those non-banking institutions that would pose a large systemic risk to the financial system, have been revisited in order to improve their ability to absorb losses and avoid a potential public bailout of these institutions.<sup>13</sup> recommends that the capital requirement should depend on the liquidity of the assets held by a bank and should increase with the proportion of its short-term debt and with its size. The Dodd–Frank Wall Street Reform and Consumer Protection Act focused primarily on the largest financial institutions in the United States.<sup>g</sup>

This paper takes the spirit of the Dodd–Frank Act to improve the credit risk aspect of the community bank. The data used in this study (loan quality rating system, term structure of the loan loss distribution, and the default rates on loans) was developed by the bank to meet the Interagency Policy Statement on the Allowance for Loan and Lease Losses dated December 21, 1993. The authors of this study empirically examine the loan quality ratings and loss distributions as well as the recovery rate (loss rate) for loans made by the community bank. The purpose code loans may have different loan losses varying from one code to another. These are related to the coordinate rating proposed by the first author Li (2014).<sup>9</sup> The coordinate rating is given by the loan grade, collateral code and purpose code as the first, second and third coordinate rating component in this case study. Individual loan may be classified from the purpose code and from the collateral code as well as the loan grading, and those different perspectives formulate the coordinate rating.

It is important to note that the loan quality rating system being studied in this paper is not the same as the credit scoring systems used by many large complex banks for small business lending on loans which, depending upon the

<sup>d</sup> Data based on the Bank Failures in Brief for FDIC-insured U.S community banks from the FDIC website. The worst year was 1992 when there were 179 banks failure in that year.

<sup>e</sup> President Barack Obama signed the Dodd–Frank Wall Street Reform and Consumer Protection Act (The Dodd–Frank Act in short) into law on July 21, 2010. The Dodd–Frank Act makes considerable progress on the off-balance sheet activities in computing capital requirements among other issues. In July 2010, Basel III is also stricter on what constitutes capital, sets a minimum leverage ratio and higher capital requirement, and creates liquidity ratios that banks will eventually have to abide by (see Dodd–Frank Act (2014) for more current Supervisory stress test methodologies and results) Even in Basel Committee (2002), banks are required to estimate the probability of default (PD) and the loss given default (LGD) on their loan portfolios of the business cycle. The data of the community bank is not enough for us to estimate these important credit risk factors.