An Efficient Data Mining for Credit Card Fraud Detection using Finger Print Recognition

[1]V.Priyadharshini, [2]G.Adiline Macriga
[2]Professor, Department of Information Technology, Chennai-44, India

Abstract

Today there are millions of credit card transactions are being processed and mining techniques are highly applied to amount transaction and processing then the data's are highly skewed Mining such massive amounts of data requires highly efficient techniques that scaled that can be extend transactions are legitimate than fraudulent fraud detection systems were widely used but this document gives the detection techniques. This paper contains multilayered techniques for providing the security for the credit card frauds. The first layer is communal detection and second is Spike detection layers that highly provides security for detection of frauds like probe resistant and mark the illegal user through their input details and mark it in a list. Then it removes attacks like defense in depths on cards and by removing the data redundancy of the attributes and it is being processed with millions of the credit cards.

Keywords

Knowledge based retrieval, Spike detection, Communal detection

1. Introduction

Fraud detection is a topic applicable to many industries including banking and financial sectors, insurance, government agencies and law enforcement, and more In banking, fraud can involve using stolen credit cards, forging checks, misleading accounting practices, etc. Identification of Identity crime is defined as broadly as possible in this project that is similar data of the user. At one extreme, synthetic identity fraud refers to the use of reasonable but fictitious identities. These are effortless to create but more difficult to apply successfully**.** At the other extreme, real identity theft refers to illegal use of innocent people's complete identity details. These can be harder to obtain (although large volumes of some identity data are widely available) but easier to successfully apply. In reality, identity crime can be committed with a mix of both synthetic and real identity details. Credit applications are Internet or paper-based forms with written requests by potential customers for credit cards.

1.1 Credit Card Fraud in Banking Industry

In the banking sector credit card fraud has increased nowadays Credit applications are Internet or paper based forms with written requests by potential customers for credit cards, loans, and personal loans. Credit application fraud is a specific case of identity crime, involving identity fraud and real identity theft then to remove duplicates in the credit card applicants. Theft rate should be reduced by Duplicates (or matches) refer to applications which share common values. There are two types of duplicates: exact (or identical) duplicates have the same values (near or approximate) duplicates have some same values (or characters), some similar values with slightly altered spellings. This paper argues that each successful credit application fraud pattern is represented by a sudden and sharp spike in duplicates within a short time, relative to the established baseline level .Duplicates are hard to avoid from fraudster view because duplicates increases their' success rate.

2. Proposed System

The main objective of this research is to achieve resilience by adding two new, real time, data mining-based layers to complement the two existing nondata mining layers proposed system utilizes real time data mining- based security layers (CD and SD) for identity crime detection. The first new layer is Communal Detection (CD): the white list-oriented approach on a fixed set of attributes. To complement and strengthen CD, the second new layer is Spike Detection (SD): the attribute-oriented approach on a variable-size set of attributes. The CD and SD layers are continuously updated. Data are traditionally based on a binary representation in which discrete information is assumed (even in continuous data, range representations are possible) and so the operations involve "modifying" bits without concern for any underlying semantics. In dealing with text data, representing the linguistic knowledge is an important issue in which traditional binary coding is insufficient, and so new representation schemes

Should be investigated.

3. Related Work

Traditional approaches to KDT share many characteristics with classical DM but they also differ in many ways: many classical DM algorithms These protocols are compatible with the existing card-based business models and payment system infrastructures. They involve three parties: the buyer (who makes the actual payment),the merchant (who will receive the payment), and the acquirer gateway (who acts as an intermediary between the electronic payment world and the existing payment infrastructure, and authorizes transactions by using the latter). Quality data are highly desirable for data mining and data quality can be improved through the real time removal of data errors (or noise). The detection system has to filter duplicates which have been reentered due to human error or for other reasons. It also needs to ignore redundant attributes which have many missing values.

4. Algorithm Implementation

CD Algorithm

1. Multi-attribute link: More applications are compared using the link types. Multi attribute link score to focus on a single link between two applications, not on matching of attributes between the values.
2. Single-link score value with average score: Average score is being created based on the user input value.
3. Parameter value change: Determine same or new parameter value by comparing inputs.
4. Whitelist creation: Valid user details are stored in the whitelist and others are rejected and new whitelist is created.

SD Algorithm

After processing the information the data are then
Send for the SD value calculation then the weights are calculated by comparing inputs.

1. Single-step scaled counts: Determine the value exceeds the time difference between each process.
2. Single value spike detection: Calculate current value score based on weighted scaled match values.
3. SD attribute selection::Each attribute weight is automatically updated at the end of the processing data stream. SD algorithm is the calculation of every current applications score using all values score and attribute weights.
4. CD attribute weights change:At the end of every current discrete data stream process,SD algorithm calculates and updates the attribute weight for CD.CD and SD is CD is provided by attribute weights by SD layer .SO both these layers together provides the single score for detection any illegal user and their scores are listed for the identity crime identification(similar values of the user)is done.

5. System Architecture and Overview

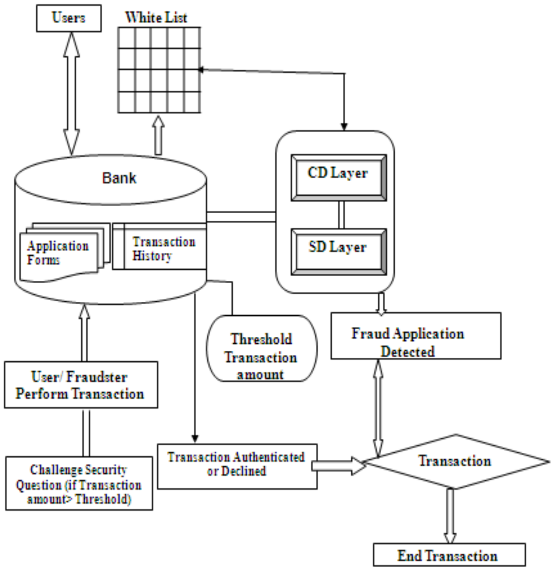
Users
White List
Bank
Application Forms
Transaction History
CD Layer
SD Layer
Threshold Transaction amount
Fraud Application Detected
User/ Fraudster Perform Transaction
Challenge Security Question (if Transaction amount> Threshold)
Transaction Authenticated or Declined
Transaction
End Transaction

Fig. 1: Architecture Diagram for Fraud Detection

6. Module Description

6.1 Communal Detection

CD layer, any two similar applications could be easily interpreted as (1) because this paper's detection methods use the similarity of the current application to all prior applications (not just known frauds) as the suspicion score. However, for this particular scenario, CD would also recognize these two applications as either (2) or (3) by lowering the suspicion score due to the higher possibility that they are legitimate. To account for legal behavior and data errors, CD is the white list-oriented between the applications, is crucial because it reduces the scores of these legal behaviors and false positives. Communal relationship is near duplicates which reflect the social relationships from tight familial bonds to casual acquaintances: family members, housemates, colleagues, neighbors, or friends broadly speaking, the white list is constructed by ranking link-types between applicants by volume. The larger the volume for a link-type, the higher the probability of a communal relationship. There are two problems with the whitelist. First; there can be focused attacks on the whitelist by fraudsters when they submit applications with synthetic communal relationships. Although it is difficult to make definitive statements that fraudsters will attempt this, it is also wrong to assume that this will not happen. The solution proposed in this paper is to make the contents of the white list become less predictable. there were two credit card applications that provided the same postal address, home phone number, and date of birth, but one stated the applicant's name to be John Smith, and the other stated the applicant's name to be Joan Smith. CD algorithm works in real time by giving scores when there are exact or similar matches between categorical data. Link score creation in current white list column.

CD algorithm is the calculation of every linked previous application's score for inclusion into the current application's score. The moving window to which the current application links. Therefore, a high score is the result of strong links between two names.

Table 1: Example for Communal Detection

Name	ID no	Street	DOB
Reena	10	S.K street	10/12/1980
Reeta	20	S.K street	10/12/1980

Table 2: This is an example for twins in a home having same address and date of birth and for this a sample white list creation is shown below from the table 1

No	Link	Score
1	00010	1

6.2 Spike Detection

The data stream point- of- view, using a series of Window steps, the SD algorithm matches the current application'svalue against varying window of previous applications' values. It calculates the current value's score by integrating all steps to find spikes. Then, it calculates the Current application's score using all values' scores and attribute weights. Also, at the end of the current Minidiscrete data stream, the SD algorithm selects the attributes for the SD suspicion score, and updates the attribute weights for CD. The ability to perform global search (traditional approaches deal with predefined patterns and restricted scope), the exploration of solutions in parallel, the robustness to cope with noisy and missing data (something critical in dealing with text information as partial text analysis techniques may lead to imprecise outcome data), and the ability to assess the goodness of the solutions as they are produced. The SD algorithm is the calculation of every current value's score by integrating all steps to find spikes. two layers, communal and spike detection, do not use external databases, but only the credit application database . And crucially, these two layers are unsupervised algorithms which are not completely dependent on known frauds but use them only for evaluation.

7. Biometrics Process for Comparison and data Extraction

Fingerprint authentication" describes the process of obtaining a digital representation of a fingerprint and comparing it to a stored digital version of a Fingerprint. Electronic fingerprint scanners capture digital "pictures" of fingerprints, either based on light reflections of the finger's ridges and valleys, ultrasonic, or the electrical properties of the finger ridges and valley. Pictures are then processed into digital templates that contain the unique extracted features of finger

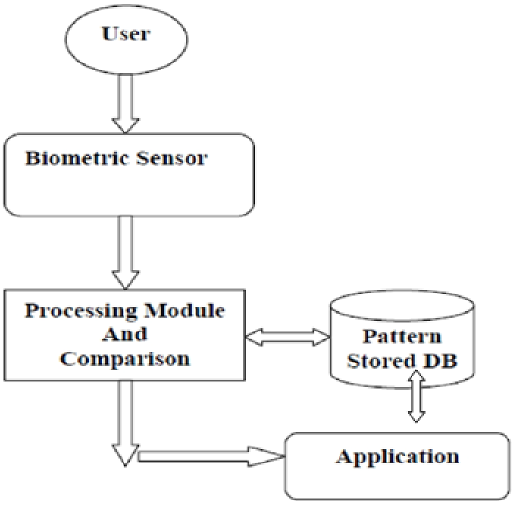

Fig. 2: Example for Biometrics Pattern Comparison and Retrieval

Calculation of suspicious score with three attributes

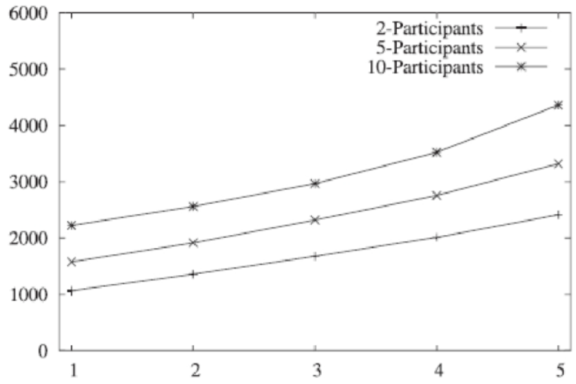

Fig .3: Graph shows that sharp spikes for

The above graph results on the data support the argument that successful credit application fraud patterns are characterized by sudden and sharp spikes in duplicates.

8. Conclusion and Future Work

This paper describes an important domain that has many problems relevant to other data mining research. It has documented the development and evaluation in the data mining layers of defense for a real-time credit application fraud detection system With the help of the finger print biometric process fraud detection can be easily found then data mining with accuracy in the existing applicants with new ones so frauds can be traced and they can be ignored without getting the card and the finger print technique is being applied in future for withdrawing the higher amounts so with these security features fraud detection can be detected.

References

[1] ID Analytics,"ID Score-Risk:Gain Greater Visibility into Individual Identity Risk,"Unpublished,2008.

[2] Feldman, Ronen, Moshe Fresko, Haym Hirsh, Yonatan Aumann, Orly Liphstat, Yonatan Schler, and Martin Rajman. "Knowledge Management: A Text Mining Approach." In PAKM, vol. 98, p. 9. 1998.

[3] Hearst, Marti A. "Untangling text data mining." In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, pp. 3-10. Association for Computational Linguistics, 1999.

[4] Aggarwal, Charu C., and S. Yu Philip. "Data mining techniques for associations, clustering and classification." In Methodologies for Knowledge Discovery and Data Mining, pp. 13-23. Springer Berlin Heidelberg, 1999.

[5] Brockett, Patrick L., Richard A. Derrig, Linda L. Golden, Arnold Levine, and Mark Alpert. "Fraud classification using principal component analysis of RIDITs." Journal of Risk and Insurance 69, no. 3 (2002): 341-371.

[6] Nahm, Un Yong, and Raymond J. Mooney. "Using information extraction to aid the discovery of prediction rules from text." In Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining (KDD-2000) Workshop on Text Mining, pp. 51-58. 2000.

[7] Christen, Peter, and Karl Goiser. "Quality and complexity measures for data linkage and deduplication." In Quality Measures in Data Mining, pp. 127-151. Springer Berlin Heidelberg, 2007.

[8] Oscherwitz, T. "Synthetic identity fraud: unseen identity challenge." Bank Security News 3, no. 7 (2005).

Dr. G.Adiline Macriga born in Nagercoil, Kanyakumari Dist on 10.01.1973.She had completed her B.Ein 1994,M.E in 2001 and Ph.D in 2011.She is an active member of ISTE.Currently she is working as professor in Department of Information Technology,Sri Sai Ram Engineering College.She had attended more than 10 National/International Confernces and published her papers in reputed Journals.

V.Priyadharshini had completed her B.E(CSE) in 2011 and presented a paper in International Conference and Currently doing P.G degree in M.E(CC) in Sri Sai Ram Engineering College.