ORIGINAL ARTICLE

Money laundering regulatory risk evaluation using Bitmap Index-based Decision Tree

Vikas Jayasree *, R.V. Siva Balan [1]

Department of Computer Application, Noorul Islam University, Kumaracoil, Thuckalay, Kanyakumari (Dt), Tamil Nadu, India

Abstract This paper proposes to evaluate the adaptability risk in money laundering using Bitmap Index-based Decision Tree (BIDT) technique. Initially, the Bitmap Index-based Decision Tree learning is used to induce the knowledge tree which helps to determine a company's money laundering risk and improve scalability. A bitmap index in BIDT is used to effectively access large banking databases. In a BIDT bitmap index, account in a table is numbered in sequence with each key value, account number and a bitmap (array of bytes) used instead of a list of row ids. Subsequently, BIDT algorithm uses the "select" query performance to apply count and bit-wise logical operations on AND. Query result coincides exactly to build a decision tree and more precisely to evaluate the adaptability risk in the money laundering operation. For the root node, the main account of the decision tree, the population frequencies are obtained by simply counting the total number of "1" in the bitmaps constructed on the attribute to predict money laundering and evaluate the risk factor rate. The experiment is conducted on factors such as regulatory risk rate, false positive rate, and risk identification time.

1. Introduction

Money laundering is the process of changing the profits of crime into apparently legal money. However, in several legal and regulatory systems, the term money laundering has developed into a combined term with other types of financial crime, and it is also used to involve misuse of the financial system.

Crime identification has become significant and extensive due to the enormous data availability on the Web and this has resulted inperpetrators preventing their original identities.

Smart Card-based Security Framework (SCSF) (Roberto Cortinas et al., 2012a,b) integrated failure detection and consensus by providing security models. The application of SCSF resulted in saving a substantial amount of messages. But the method did not focus on money laundering based security failure detection. Multilayered Detection System (MDS) (Phua et al., 2012) was designed to determine the credit money laundering fraud detection by applying Communal Detection (CD) and Spike Detection (SD). However, effectiveness with respect to scalability remained unsolved.

Effective identification of patterns can be performed to measure the audit work and avoid tax evasion. The clustering

* Corresponding author at: Global IT Department, Habib Bank AG Zurich, Post Box 3306, Dubai, United Arab Emirates. Tel.: +971 50 503 5529.
E-mail addresses: jvikas77@gmail.com (V. Jayasree), rvsivan@gmail.com (R.V. Siva Balan).
[1] Tel.: +91 9445245689.

algorithms (Castellón González and Velásquez, 2013) like SOM and neural gas were applied for the effective identification of group behavior and accordingly the patterns were detected to identify the fraudulent activities. Another method, anti money laundering (AML) solutions (Eldin Helmy et al., 2014) were designed with the objective of improving the detection process using rule based monitoring, behavioral and clustering monitoring. However, the retrieval of the detection process reduced with the increased user accounts.

A method called, bitmap index (BI) (Laxmaiah et al., 2013a,b) was designed with the motive of fast retrieval of information using a priority queue (PQ) and Ice Berg (IB) queries. The method was proved to be efficient in terms of execution time with different thresholds. However, reducing redundant bits remained unaddressed. To solve this issue, data mining techniques with Bit Map Vectors (BMV) (Laxmaiah et al., 2013a,b) were applied to minimize the redundant bits by introducing compacted number of AND operations. Fraud detection methods using Neural Network and Support Vector Machine (Zareapoo et al., 2012) were constructed to reduce fraud detection in credit card applications.

In this paper, we design a Bitmap Index-based Decision Tree for risk evaluation on financial money laundering to improve the adaptability rate using a technique called Bitmap Index-based Decision Tree (BIDT). The contributions of BIDT include the following:

- To evaluate the adaptability risk in money laundering using Bitmap Index-based Decision Tree (BIDT) technique.
- To efficiently determine the company's money laundering risk and improve the scalability using Bitmap Index-based Decision Trees learning.
- To significantly evaluate the adaptability risk in money laundering operation by constructing a decision tree in a more precise manner.
- To extensively predict the money laundering and evaluate the risk factor rate by obtaining the population frequencies by simply counting the total number of integers in the bitmaps.

The structure of the paper is as follows. Section 2 discusses the related works on money laundering by various researchers. Section 3 describes in detail the proposed technique, Bitmap Index-based Decision Tree. Section 4 provides the experimental setup and Section 5 discusses in detail the parameter definitions with the help of table and graph form. Finally, Section 6 concludes with concluding remarks.

2. Related works

The recent crackdown in financial institutions has resulted in the largest form of cross-border money laundering that has become one of the greatest potential vehicles for money laundering. The reverse engineering methods (Möser et al., 2013) were applied to study the opportunities and limitations related to anti-money laundering. Another method called, money laundering under electronic payment (Weibing, 2011) was designed with the motive of preventing electronic money laundering crime and possibility of early detection (Pulakkazhy and Balan, 2013). However, the financial computerization tendency remained unsolved.

With the increasing advancement and development in the field of Internet money laundering in banks, the volume of data is growing at a faster speed. An algorithm based on classification (Luo, 2014) was designed for the effective identification and detection of suspicious activities (Jayasree and Siva Balan, 2015). Though suspicious activities were detected at an earlier stage, it was done at the cost of time. Indexing techniques (Suresh and Thammi Reddy, 2014) were used to identify the relationship between attributes and graph theoretic approach was applied with the aid of apriori algorithm to reduce the retrieval time.

One of the main problems when dealing with money laundering is the handling of voluminous financial information. The main drawbacks are due to the fact that there appears no identical tactics with financial institutions to deal with the anti-money laundering. Digital forensics and database analysis were integrated (Flores et al., 2011) to verify the customer in an extensive manner and detect fraudulent activities (Jayasree and Siva Balan, 2013). However, still it has challenged the evaluation of the digital forensic practices within the organization. Though legalization is very difficult to apply, the harmonization approach (Nikoloska and Simonovski, 2012) was applied for the early detection of crime. Adaptive join operators (Bornea et al., 2010) resulted in the optimization of the customer activities through multiple index nested loop reactive join.

Based on the aforementioned methods, in this work, we propose a technique called Bitmap Index-based Decision Tree for effective evaluation of risk factor on financial money laundering in banks.

3. Bitmap Index-based Decision Tree for risk evaluation on financial money laundering

The main objective of the proposed work is to evaluate the risk factor on financial organizations money laundering using the indexing scheme. The indexing scheme uses the rows and columns to store the information to improve the scalability rate. Money laundering is the illegal amount transacted between different users, which are evaluated using the Bitmap Index-based Decision Tree (BIDT) technique. The risk related to the larger amount of illegal transaction is controlled in a financial organization by constructing a decision tree with mapping of the bit in fuzzy form '0' and '1'. The decision tree contains the root and sub co-ordinate nodes to create the determination rules in the BIDT technique. The purpose of indexing in BIDT technique is to provide pointers to the rows in a table containing given key values.

Fig. 1 shows the bitmap structure representation. The money laundering account is effectively evaluated with bitmaps using cardinality rows and columns. BIDT technique effectively relates the data in the database table for easy evaluation. The bitmap index uses the arrays and bitwise logical operation AND for easy evaluation of results. In the figure, the horizontal distributed black dots indicate the row ID list and vertically arranged dots indicates the column ID list. From the figure, while considering the account number is 1001, the row id list is calculated according to the three horizontal black dots. Similarly, the black dots are distributed along with the column id list measured in a perpendicular way. The bitmap is indexed through the column Id list and row Id list for easier
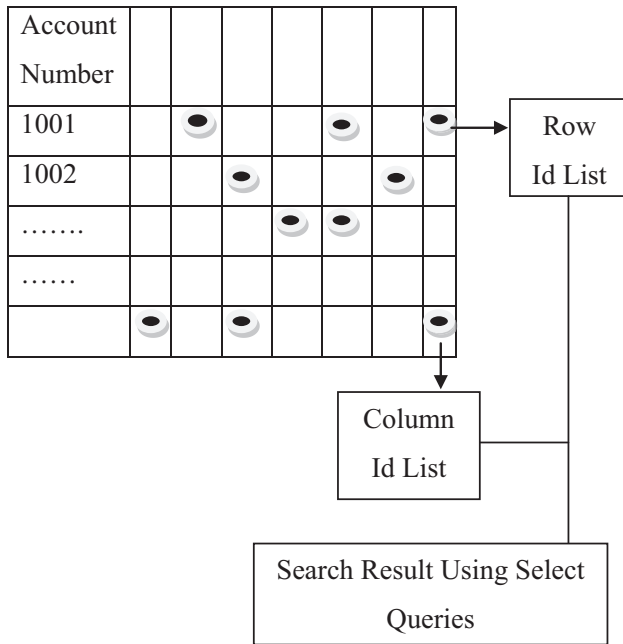
Figure 1   Bitmap structure representation.

result fetching in BIDT technique. Bitmap indexing is a significant system in the proposed work, with the performance advantage over the system. For instance, the bitmap indexing is used to easily evaluate money laundering on the online transaction processing in financial organizations. The online transaction risk evaluation is a critical task in the existing system. The architecture diagram of Bitmap Index-based Decision Tree (BIDT) technique is shown in Fig. 2 for easy description of the risk evaluation.

As illustrated in Fig. 2, the account details of money laundering are noticed and then the bitmap indexing procedure is applied. The bitmap indexing uses the rows and column id information on the large banking database. The two steps such as indexing with the help of the row and column table and decision tree are carried out in the design and implementation of BIDT technique. Indexing in BIDT is carried out using the select query option and logical AND operator. These results are used to construct the decision tree, where the customer region of transaction and risk occurrence are identified. Money laundering activities occurring with high frequency are measured and processing is carried out with higher efficiency.

3.1. Bitmap indexing

In BIDT technique, bitmap index is a data structure which is used to efficiently access larger bank databases involving money laundering accounts. The proposed index work is to provide pointers to the rows in the table where bulk transaction is carried out. The risk factor is analyzed easily using the given distinct key value on each transaction for a particular account. In the common bitmap index, list of row id and column id are used to identify whether a particular bulk transaction is carried out on the specific customer account. The map index contains the record in the table in a sequential fashion.

Bitmap indexing in the proposed work produces the result with the fuzzy form (i.e., 0 or 1). For each distinct key value,
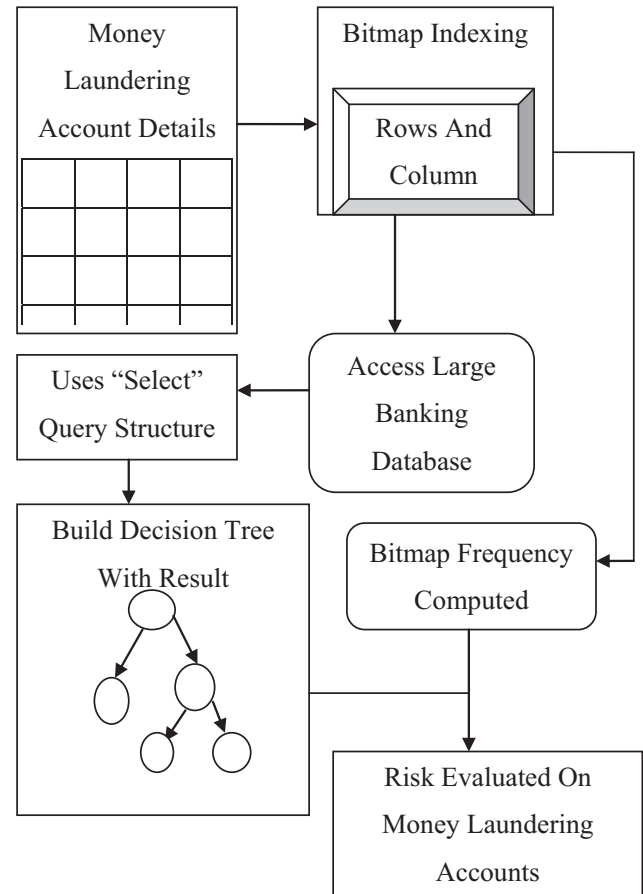

Figure 2   Architecture diagram of BIDT technique.

bit mapping is used for easy risk evaluation on money laundering accounts based on the fuzzy form. Each bit in the bitmap corresponds to the existing row and column id during the mapping operation. All the rows and columns in the bitmap indexing have distinct values for easy evaluation of the risk. Bitmap indexing method efficiently categorizes the rows and columns based on the account details of the customer that reduce the risk evaluation time, since the usage of the key value in BIDT technique.

3.1.1. Select query structure

Bitmap indexing is used for efficient query based risk evaluation on money laundering account with multiple different key value databases. The Query is answered using bit wise logical operator and it is formularized as,

Select Bank Table where customer transactionis >>
    specified range AND count the transaction time   (1)

Each operation takes bitmap indexing with the same size of customer account and computes the desired tuple range. SQL queries are used on the bitmap indexing table in the BIDT technique to efficiently perform special operations using combinations of multiple indices. The select query structure in BIDT technique is used to extensively provide customer satisfaction. The logical 'AND' operator in the BIDT technique helps to easily combine the queries and produce the desired result.

3.1.2. Bitmap index frequency

Bitmap index frequency depends on the special type of indexing using the row and column id. The computation of frequency is easier as lower cardinality values are used during the computation. The advantage of using a Low cardinality column in the BIDT technique is that it has a unique key value for easier evaluation of the risk on the money laundering accounts. Bulk transaction amount of the accounts are identified and risk occurred in the transaction is analyzed in the BIDT technique. The bitmap frequency using low cardinality column is computed as,

$$\text{Frequency} = \sum_{i=1}^{n} A_c[\pi_i, \pi_{i+1} \ldots \pi_{i+n}] \qquad (2)$$

From (2), the frequency point is measured which is the overall sum of the attributes for analyzing the risk (i.e., type of attributes used to identify the customer way of the transaction) factor. In (2), A_c denote the attribute column in the BIDT technique in which π_i is the account used for each transaction. Money laundering is carried out on different intermediate accounts $\pi_{i+1}, \pi_{i+2} \ldots \pi_{i+n}$ to reach the particular customer without any uncertainty. Therefore, money laundering consideration with low cardinality is carried out in the BIDT technique to improve the regulatory risk rate.

3.2. Decision tree structure

Decision tree structure is the most popularly used data mining technique to analyze the risk factor. The decision tree structure in the BIDT technique is used to partition the decision into smaller partitions for analyzing the money laundering factor. BIDT technique takes the input as the set of objects (i.e., customer accounts) with the predictive attributes 'A'. The location, business type, age, gender are other three sequential important attributes. For each other node (i.e.,) the intermediate accounts in the decision tree, new set of bitmaps are generated, each one corresponding to the class in the node. The diagrammatic form of the decision tree is shown in Fig. 3.

The results of the queries coincide exactly to build a decision tree and more precisely to evaluate the adaptability risk in the money laundering operation. The attribute predicts the money laundering and evaluates the risk factor rate. The Bitmap indices (i.e., 0 and 1) improve the performance by applying count and bit-wise logical operations AND. To obtain the root node of the decision tree, the population frequency of each class is to be determined. For the root node (i.e.,) main account of the decision tree, the population frequencies are obtained by merely including the total number of "1" in the bitmaps. The number of '1' in the bitmap denotes the transaction carried out between the intermediate accounts. The main and intermediate account level of money transaction is noted and it is partitioned as low, medium and high range.

Each account belongs to the set of the mutually exclusive attribute classes. Decision tree construction applies successive criteria to obtain the partition and produce the effective risk factor evaluation rate on the money laundering. The risk factor is analyzed and decided effectively using the Advanced Iterative Dichotomiser 3 (AID3) algorithm. AID3 algorithm partitions the node with entropy and the algorithmic step wise description is briefly provided in Section 3.2.1.
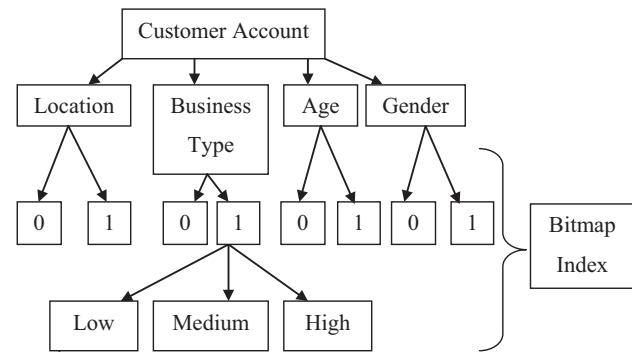
Figure 3 Decision tree based representation of the account details for risk evaluation.

3.2.1. Tree structure step wise description

AID 3 is used widely for generation of the decision tree for money laundering risk evaluation. BIDT technique uses the Occam's razor principle, where all the attributes are used for the entropy computation. The processing step is described as,

Begin
Step 1: Senses all the attributes in the account
Step 2: For Each Attribute 'A'
    Step 2.1: Entropy is computed with the maximum information gain

$$\text{Entropy} = \sum_{i=1}^{n} \log_2 P_i, \; P_i \text{ is the positive value range}$$

Step 3: For every positive value on transaction
    Step 3.1: Add the new tree branch below the root to identify the bitmap index range
    Step 3.2: Bitmap index with the '1' is analyzed and level of the transaction is noticed
    Step 3.3: Complete tree till the leaf node with the target value range
Step 4: End For
Step 5: Entropy achieves the positive and negative value
Step 6: Negative values are discarded, by minimizing the risk evaluation time
Step 7: Finding leaf data reduced the number of test on pruning (i.e., risk evaluation)
Step 8: End For Each
End

The prediction rule in AID3 helps to easily analyze the risk adaptability rate. The Occam's razor principle is used for obtaining all the attributes in the financial banking accounts. The accounts which show money laundering with high risk factor are easily analyzed with the bitmap indexing procedure. The indexing with the key values and decision tree construction attains the highly balanced result set.

4. Experimental evaluation

Bitmap Index-based Decision Tree (BIDT) Technique is implemented for evaluating the money laundering risk in the financial institutions. BIDT effectively performs the evaluation

using the JAVA platform. Statlog German Credit Data from the UCI repository classifies the people by a set of attribute lists for the easier evaluation of the risk factor. Indexing technique is the effective way to evaluate the risk factor rate using the Strathclyde University information. The numerous pointer variables are added to make an effective algorithm for money laundering risk evaluation. Data mining has the capability to evaluate the risk of money laundering, fraud system because these techniques use effective models.

The attribute characteristics are categorized as categorical and integer type. Nearly, 17 attributes from the Statlog German Credit Data have been coded as integer type and 3 under the categorical type. Bitmap Index-based Decision Tree (BIDT) Technique is compared with the existing method such as TrustedPals, a Smart Card-based Security Framework (SCSF) (Roberto Cortinas et al., 2012a,b) and Multilayered Detection System (MDS) (Phua et al., 2012). Experiment is conducted on the factors such as regulatory risk rate, false positive rate, Adaptability rate, and risk identification time.

5. Discussion

The result analysis of BIDT technique using Statlog German Credit Data is compared with existing Smart Card-based Security Framework (SCSF) (Roberto Cortinas et al., 2012a,b) and Multilayered Detection System (MDS) (Phua et al., 2012).

5.1. Impact of risk identification time

Risk identification time in BIDT technique is the time taken to identify the key values on money laundering accounts in the bank. Higher the time to identify the key values, the higher the risk identification time and vice versa. It is measured in terms of milliseconds (ms).

$$RI_{\text{time}} = \sum_{i=1}^{n} \frac{K_i}{\text{Size of user account}} \qquad (3)$$

Fig. 4 shows the impact of risk identification time with user account of size 50–350 KB and comparison made with two other methods SCSF and MDS respectively to that of the proposed method BIDT technique. From the figure it is evident that the risk identification time is reduced using the proposed BIDT technique when compared to SCSF (Roberto Cortinas et al., 2012a,b) and MDS method (Phua et al., 2012). The risk identification time in BIDT technique is reduced with the application of bitmap based indexing. With the application of bitmap based indexing, during mapping operation, BIDT uses the fuzzy form where bitmap corresponds to the existing row and column id. Using the fuzzy form, the rows and columns have distinct values which help in reducing the risk identification time using BIDT technique by 9–25% compared to SCSF and to 13–60% compared to MDS respectively.

5.2. Impact of false positive rate

The false positive rate in BIDT technique measures the transaction associated with a genuine customer who is blocked because of a name or another match. Lower the false positive rate, more efficient the method is said to be. It refers to the genuine people incorrectly identified as not genuine. It is
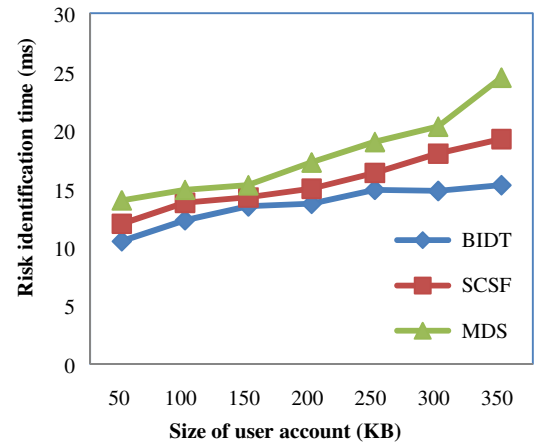

Figure 4 Measure of risk identification time.

measured in terms of percentage (%) and is formalized as given below

$$FPR = \frac{\text{Gen}_{\text{incorrectly identified as not genuine}}}{\text{User counts}} \qquad (4)$$

The comparison of false positive rate is presented in Table 1 with respect to the number of user counts in the range of 50–350. With an increase in the number of user counts, the false positive rate also gets increased, though not linear due to the variation in genuine cases and accordingly the false positive rate also varies.

To ascertain the performance of the false positive rate, comparison is made with two other existing methods Smart Card-based Security Framework (SCSF) (Roberto Cortinas et al., 2012a,b) and MDS method (Phua et al., 2012). In Fig. 5, the number of user counts for experimental purpose varies between 50 and 350. From the figure it is illustrative that the false positive rate is reduced using the proposed BIDT technique when compared to the two other existing methods. The false positive rate in the BIDT technique is significantly reduced by applying the select query structure. The select query structure uses multiple different key value databases on money laundering account and computes the desired tuple range. The identification of desired tuple, range in BIDT technique helps in reducing the false positive rate by 7–26% compared to SCSF. Moreover, special operations are performed using combinations of multiple indices using logical 'AND' operator in the BIDT technique that helps to easily combine the queries and produce the desired result and therefore minimizing the false positive rate by 19–55% compared to MDS.
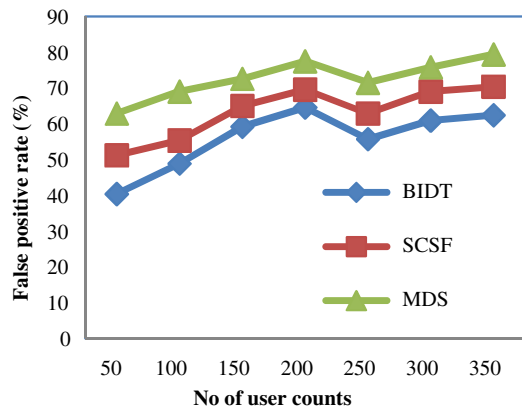
5.3. Impact of true positive rate

True positive rate refers to the measure of genuine customer correctly identified as genuine. It is measured in terms of percentage (%). Higher the true positive rate, the more efficient the method is. True positive rate is formalized as given below

$$TPR = \frac{\text{Gen}_{\text{correctly identified as genuine}}}{\text{User counts}} \qquad (5)$$

In Fig. 6, the true positive rate using the number of user counts of size 50–350 for experimental purposes is depicted. From the figure, the value of the true positive rate achieved

Table 1 Tabulation for false positive rate.

No of user counts	False positive rate (%)		
	BIDT	SCSF	MDS
50	40.35	51.19	62.86
100	48.85	55.34	68.98
150	59.19	64.98	72.56
200	64.55	69.55	77.45
250	55.69	62.85	71.44
300	60.89	68.98	75.79
350	62.38	70.35	79.35

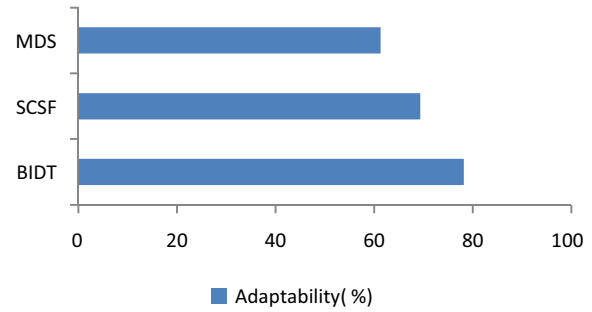**Figure 7** Measure of adaptability.

Figure 5 Measure of false positive rate.
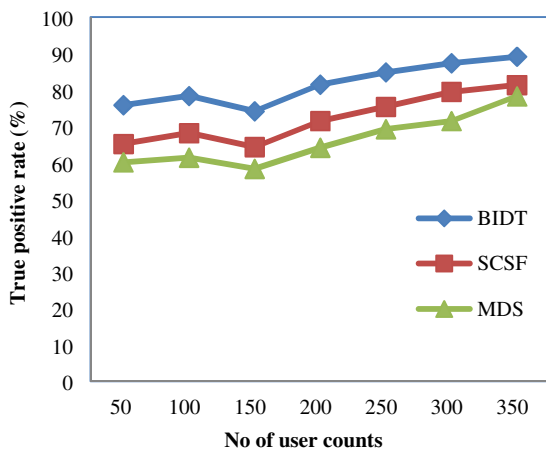
Figure 6 Measure of true positive rate.

using the proposed BIDT technique is higher when compared to two other existing methods SCSF (Roberto Cortinas et al., 2012a,b) and MDS method (Phua et al., 2012). Besides, we can also observe that by increasing the size of the user counts, the value of the true positive rate is increased using all the methods. But comparison, it is higher in the BIDT technique. The true positive rate (i.e., regulatory risk rate) is improved on the BIDT technique by evaluating the bitmap index frequency. By evaluating the bitmap index frequency where lower cardinality values are used that has unique key value results in

the improvement of the true positive rate using BIDT technique by 8–14% compared to SCSF. In addition, a bulk transaction amount of accounts is identified with the aid of frequency point that represents the overall sum of the attributes on each transaction. This in turn improves the true positive rate in BIDT by 12–21% compared to MDS.

5.4. Impact of adaptability rate

The adaptability rate of crime using the BIDT technique is the ability of the service provider (i.e., bank) to adjust changes in services based on customers' requests during money laundering operation. Adaptability on crime measures the time taken to adapt to changes or update the money laundering service to a higher level at a less interval of time. Higher the adaptability rate, more quickly, the anti money laundering system is and therefore is said to be more efficient in handling the money laundering operations.

Fig. 7 shows the measure of adaptability using the BIDT technique, SCSF and MDS respectively. From the figure it is illustrative that the adaptability rate using the BIDT technique is higher when compared to the existing methods SCSF and MDS. The rate of adaptability in the BIDT technique improves with the application of decision tree structure that efficiently partitions the decision into smaller partitions. Furthermore, with the application of Advanced Iterative Dichotomiser 3 (AID3) algorithm, that effectively partitions the nodes with entropy through Occam's razor principle where every attribute is used for entropy evaluation, improving the rate of adaptability by 11.23% and 11.67% compared to SCSF (Roberto Cortinas et al., 2012a,b) and MDS method (Phua et al., 2012) respectively.

6. Conclusion

Maintaining regulatory risk rate and providing security for financial organizations has become the key for money laundering with the main objective of improving the level of the true positive rate (i.e., regulatory risk rate) by reducing the time taken in identifying the risk. In this work, the performance effects of regulatory risk evaluation are investigated called as Bitmap Index-based Decision Trees. Bitmap indexing method efficiently categorizes the rows and columns based on the account details of the customer that reduces the risk identification time and greatly improves the adaptability rate. First, we study the use of Bitmap Indexing that efficiently handles large

money laundering accounts and produces the result with a fuzzy form to improve the regulatory risk rate. Second, we develop Select Query Structure with multiple key value databases that work with the bitwise logical operator to minimize the false positive rate. We also integrate Bitmap Index Frequency with the rows and columns id using the Low Cardinality column for improving the true positive rate using Statlog German Credit Data from the UCI repository. The experiment conducted using Statlog German Credit Data from the UCI repository shows that the BIDT technique outperforms in terms of the true positive rate, false positive rate, the risk identification time and adaptability rate when compared to the state-of-the-art methods.

Conflict of interest

No conflict of interest.

References

Bornea, Mihaela A., Vasilis, Vassalos, Yannis, Kotidis, 2010. Adaptive join operators for result rate optimization on streaming inputs. In: IEEE Trans. Knowl. Data Eng. 22 (8).

Castellón González, Pamela, Velásquez, Juan D., 2013. Characterization and detection of taxpayers with false invoices using data mining techniques. Expert Syst. Appl. (Elsevier)

Eldin Helmy, Tamer Hossam et al, 2014. Design of a monitor for detecting money laundering and terrorist financing. Int. J. Comput. Networks Appl. 1 (1).

Flores, Denys A., Angelopoulou, Olga, Self, Richard J., 2011. An anti-money laundering methodology: financial regulations, information security and digital forensics working together. J. Internet Serv. Inf. Secur. (JISIS) 3.

Jayasree, Vikas, Siva Balan, R.V., 2013. A review on data mining in banking sector. In: Am. J. Appl. Sci. 10 (10), 1160–1165.

Jayasree, Vikas, Siva Balan, R.V., 2015. Money laundering identification on banking data using probabilistic relational audit sequential pattern. Asian J. Appl. Sci., 1996–3343

Laxmaiah, M. et al, 2013a. A compressed bitmap vector method to assess aggregate queries competently. Int. J. Adv. Res. Comput. Sci. Software Eng. 3 (11).

Laxmaiah, M. et al, 2013b. An approach to evaluate aggregate queries efficiently using priority queue technique. Int. J. Emerging Trends Technol. Comput. Sci. 2 (3).

Luo, Xingrong, 2014. Suspicious transaction detection for anti-money laundering. In: Int. J. Secur. Appl. 8 (2).

Möser, Malte, Böhme, Rainer, Breuker, Dominic, 2013. An inquiry into money laundering tools in the bitcoin ecosystem. eCrime Researchers Summit, IEEE.

Nikoloska, Svetlana, Simonovski, Ivica, 2012. Role of banks as entity in the system for prevention of money laundering in the Macedonia. Social Behav. Sci. (Elsevier)

Phua, Clifton et al, 2012. Resilient identity crime detection. IEEE Trans. Knowl. Data Eng. 24 (3).

Pulakkazhy, Sreekumar, Balan, R.V.S., 2013. Data mining in banking and its applications-a review. J. Comput. Sci. 9 (10), 1252–1259.

Roberto Cortinas et al, 2012a. Secure failure detection and consensus in trusted pals. IEEE Trans. Dependable Secure Comput. 9 (4).

Roberto Cortinas et al, 2012b. Secure failure detection and consensus in trusted pals. IEEE Trans. Dependable (n.d.).

Suresh, C.H., Thammi Reddy, K., 2014. Graph based approach to identify suspicious accounts in the layering stage of money laundering. Global J. Comput. Sci. Inf. Technol. 1 (1), 81–87.

Weibing, Peng, 2011. Research on money laundering crime under electronic payment background. J. Comput. 6 (1).

Zareapoo, Masoumeh, Sreeja, K.R., Afshar Alam, M., 2012. Analysis of credit card fraud detection techniques: based on certain design criteria. Int. J. Comput. Appl. 52 (3).