International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 7– No.9, September 2014 – www.ijais.org

A Novel Method to Detect False Financial Statement using Negative Selection Algorithm

U.Jothi Lakshmi
AP/Department of IT
PITAM, Thiruvallur, Tamilnadu,
India

ABSTRACT
Financial statement fraud is one of the biggest challenges in the modern business world. It affects various sectors of people including the fraudsters, auditor and the public. Above all the economic growth of a country diminishes adversely. So the need to prevent such fraud is very important. But as the fraudsters are so adaptive to new trends it is hard to develop a preventive mechanism. And the job of auditors is very much time consuming that the chance of misinterpretation is also high in nature. Hence this paper proposes a detection mechanism - that include artificial immune algorithm - is supposed to be capable of detecting false financial statement effectively.

General Terms
Fraud Detection

Keywords
FFS (False Financial Statement), Negative Selection Algorithm, Artificial Immune System.

1. INTRODUCTION
Financial statement is an on paper report which explains the fiscal strength of a company [2]. It is a quantitative report which includes, Balance sheet, Statement of income or statement of operations, Statement of retained earnings, Statement of cash flows, Statement of changes in owner's equity, Notes. General definition of FFS is intentional wrong reports or exclusion of amounts or disclosure of fiscal reports to misinform financial statement users, particularly investors and creditors [1]. Financial statement fraud can mature in several unusual forms, however once deceptive accounting practices are initiated; various systems of manipulation will be utilized to maintain the appearance of sustainability. Common approaches to artificially improving the appearance of the financials include: overstating revenues by recording future expected sales, understating expenses through such means as capitalizing operating expenses, inflating assets' net worth by knowingly failing to apply an appropriate depreciation schedule, hiding obligations off of the company's balance sheet and incorrect disclosure of related party transactions and structured finance deals. Another alternative to financial statement fraud involves cookie-jar accounting practices, a procedure by which a firm will understate revenues in one accounting period and maintain them as a reserve for future periods with worse performance. Such procedures remove the appearance of volatility from the operations.

These types of practices would affect various people including the fraudster himself/herself. In addition to that the economic growth of the country goes down remarkably. The job auditors, who are responsible in identifying truth/false behind a financial statement is a tedious one and also time consuming. An experienced auditor can identify FFS effectively when compared to the novice one. But still human errors are unavoidable. Hence there is a strong need for an effective mechanism.

The Negative Selection Algorithm (NSA) is a strategy inspired from biological immune system. In biological immune system, negative selection plays a key role. In that the system learns its own cells called self-cells. Based on the self-cell it develops antibodies that match with foreign cell i.e., non-self-cell but not with self-cell. If a match is found then the antibody engulfs the non-self and kills it. If an antibody identifies self-cell as non self then the effect is auto immunity. Auto immunity is dangerous one and a fatal symptom too.

This paper proposes a potentially effective system that uses NSA in FFS detection. Here the genuine statements are considered as self and fraudulent one as non-self. The paper is organized as follows, Section 1 Introduction, Section 2 Literature Review, Section 3 System Structure, Section 4 Conclusion and Section 5 References.

2. LITERATURE REVIEW
In this e world things are happening very fast and quick. Making money by cheating has become a business worldwide. Many organizations worldwide give false financial statement so as to project themselves low financially and gain more profit. In some other cases few project themselves high and get profit; like getting loan or assets or anything that is owned by other entity. All such thing happens because of well-formed false financial statement. So it is obvious now a days to detect whether a given financial statement is genuine or not. The following section gives you works that were done on Financial Fraud Detection done using data mining techniques which is followed by description about Negative Selection Algorithm.

2.1 Financial Fraud Detection
In [4], E.W.T. Ngai et.al have done research and reviewed various literatures on how to apply data mining technique to detect fraudulent financial statement. The paper gives an extensive study on various Financial Fraud Detection (FFD) methods. They concluded that the research on FFD is still lacking. In [5] Wei Zhou et.al., have considered , classification algorithm, Regression, and neural networks as the techniques that play significant role in detecting Financial Statement Fraud (FSF). In [3] P. Ravisankar et.al., have used financial statement dataset for their research. This dataset comprises instances of 202 Chinese companies that are expressed using 35 attributes as listed in Table 1. Being a high dimensional dataset it eventually leads to curse of

dimensionality which affects the performances of any data mining algorithm especially the classifiers. Hence the need of feature selection is insisted in this paper. They have analyzed the algorithms, MLFF, SVM, GMDH, GP, LR and PNN with all 35 features, selected 18 in first case and selected 10 features in second case. They have used t-statistic for feature selection method to reduce it to 18 in first case and to 10 in second cases. They have proved that the reduced feature subset have improved the performance of MLFF, SVM, GMDH, GP, LR and PNN. In [6], Fletcher H. Glancy et.al, have taken two vital stream for their research, 1. Research into corporate finance reporting fraud, 2. Text – mining research. They have done their research by understanding the theories behind fraud detection and applying text mining in fraud detection. They have given a Computational Fraud Detection Model (CFDM) which uses text mining strategy to detect fraudulent financial statement.

Table 1. Attributes from financial statements [3]

Sl.No.	Financial attributes
1	Debt
2	Total assets
3	Gross profit
4	Net profit
5	Primary business income
6	Cash and deposits
7	Accounts receivable
8	Inventory/Primary business income
9	Inventory/Total assets
10	Gross profit/Total assets
11	Net profit/Total assets
12	Current assets/Total assets
13	Net profit/Primary business income
14	Accounts receivable/Primary business income
15	Primary business income/Total assets
16	Current assets/Current liabilities
17	Primary business income/Fixed assets
18	Cash/Total assets
19	Inventory/Current liabilities
20	Total debt/Total equity
21	Long term debt/Total assets
22	Net profit/Gross profit
23	Total debt/Total assets
24	Total assets/Capital and reserves
25	Long term debt/Total capital and reserves
26	Fixed assets/Total assets
27	Deposits and cash/Current assets
28	Capitals and reserves/Total debt
29	Accounts receivable/Total assets
30	Gross profit/Primary business profit
31	Undistributed profit/Net profit
32	Primary business profit/Primary business profit of last year
33	Primary business income/Last year's primary business income
34	Account receivable /Accounts receivable of last year
35	Total assets/Total assets of last year

2.2 Negative Selection Algorithm

Artificial Immune System is a computational model inspired by natural immune system [9]. Negative Selection is a mechanism which belongs to Immune System, which deals with self and non-self-cells. The negative selection algorithm was first used by Forrest et al [8] as a means of identifying suspicious strings for virus detection in computer systems. Fig.1 [9], shows the flow of Negative Selection Algorithm. It has two vital phases. Phase 1: Generation potential Detector set, Phase 2. Classify /Detect new Instance using potential Detector Set. In phase 1 the input is randomly generated sample. These samples could be represented as binary or as a real value [7].

Phase 1: Generating potential Detector set

In this phase, known genuine sample of data is considered as self. Random instances/candidates are generated and compared with self. If a match is found then it is rejected, if not it is preserved and called as simply detectors or potential detector see fig.2.

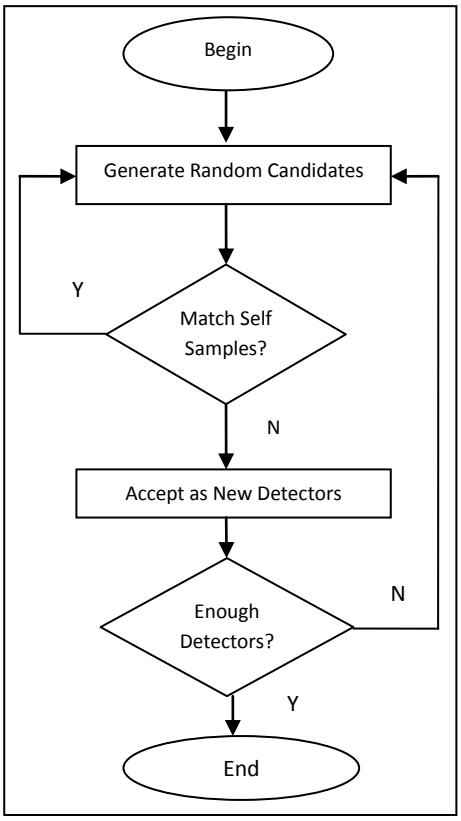
Phase 2: Classifying/Detecting new Instance using potential Detector Set

In this phase, each preserved potential detector is compared with new unknown instance of data or vice versa. If a match is found then the given instance is labeled as non self see Fig.3.

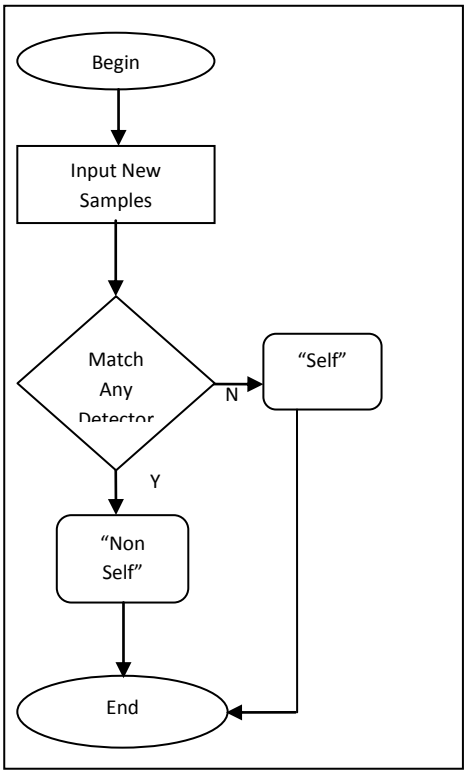
3. SYSTEM STRUCTURE

This paper proposes a new method to detect false financial statement using Negative Selection Algorithm (NSA). It leverages the benefits of NSA. The initial stage of this proposed system structure is to collect financial statements (FS) of all categories from reputed agencies which deal with auditing the FS.

The Detector Generation Engine (DGE) specified in Fig. 4 is an automated system that takes all FS as input and by itself categorizes it as genuine or false. The collected raw data could be converted to a dataset with 35 attributes as given in Table 1.



(a)Phase1: Generating potential Detector set



(b)Phase2: Classifying/Detecting new Instance using potential Detector Set

Fig 1: Outline of typical Negative Selection Algorithm [9]

Fig. 2 Sample illustration of Phase 1 if SELF is given as input

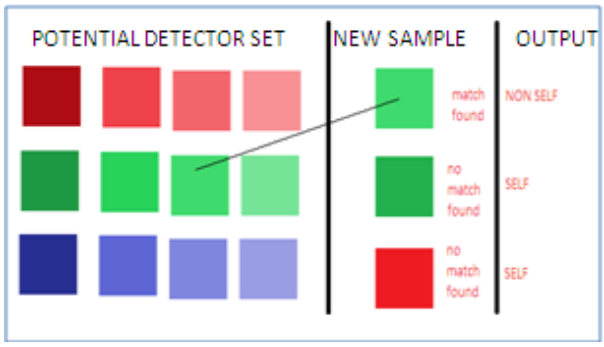


Fig. 3 Sample illustration of Phase 2. Three input samples given as input

This paper doesn't explains much about this part as it is very extensive procedure and implementation of the proposed system isn't started yet. Instead readily available dataset could be used for the discussions about the proposed system. The data set is segregated into two categories. 1. All genuine instances (AGI) , 2. All false instances (AFI). These two categories of instances play a vital role in the proposed structure. Randomly generated values by NSA are compared with AGI, if match is found it is ignored otherwise preserved as Detectors. Detector is then compared with AFI if match is found then it is preserved as Effective Detector Set (EDS), it is also called as memory cell as it is similar to already occurred event. Otherwise it is preserved under Auxiliary Detector Set (ADS). The term memory cell is not analogous to the same term in Clonal Selection Algorithm another unit of Artificial Immune System. The comparison done by the test module in DGE is based on matching rule. The matching is uses a threshold value (T.value). If the match found exceeds or equals T.value then Detectors are moved to EDS else to ADS.

Now FFS detection system acquires set of detectors from Detector Generation Engine (DGE) then compares it with new instance given for verification. The system provides a result as Genuine if no match found else as False. The proposed system has higher potential to provide better result in classifying the given statement as genuine or false. The reason behind this is EDS and ADS. It is like having a readymade copy of false statement and the systems job is make a comparison between them. Here the role of DGE is very crucial. And finally the role of Administrator is to provide input and view the result.

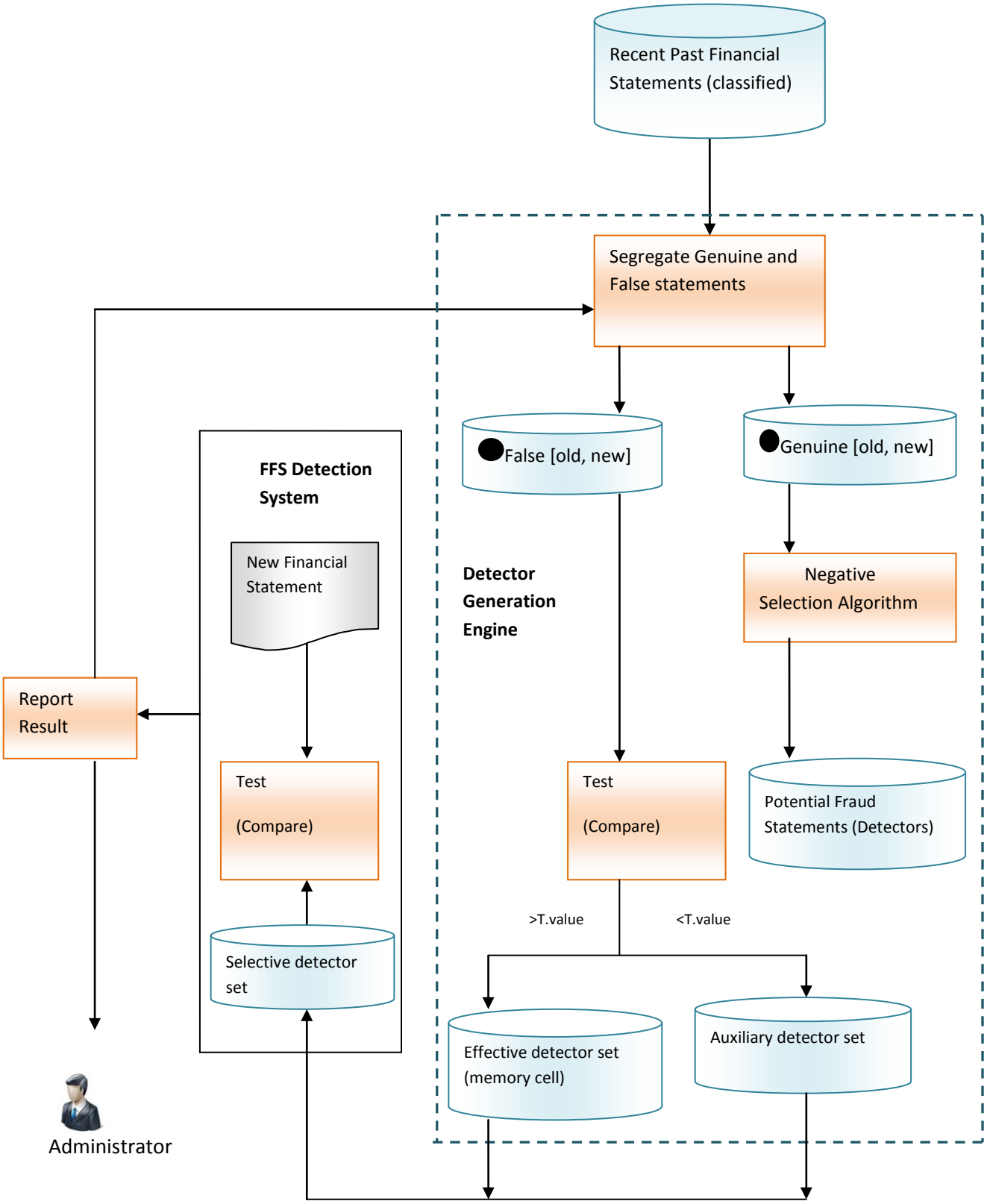


Fig. 4. System structure of False Financial Statement Detection using NSA

4. CONCLUSION

This paper proposes a system structure to detect false financial statement using a bio inspired classification procedure called Negative Selection Algorithm. The proposed system has a lot of planning and preparation stages. The ultimate aim of the proposed system is to bring up a chunk of good EDS and ADS. If this becomes feasible then it eases the job of detecting the fraudulent statements. The challenges in bringing up the proposed system are 1. Representation of the instances. 2. Selecting relevant features, 3. Identifying number of detectors to be preserved. If the above mentioned challenges are approached with right methodology, the proposed system would solve the issues that arise due to false financial documents.

5. REFERENCES

[1] www.passia.org/

[2] http://www.investorwords.com/

[3] Detection of financial statement fraud and feature selection using data mining techniques P. Ravisankar , V. Ravi , G. Raghava Rao , I. Bose , Decision Support Systems 50 (2011) 491–500,Elsievier.

[4] The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, E.W.T. Ngai , Yong Hu ,Y.H. Wong , Yijun Chen , Xin Sun , Decision Support Systems 50 (2011) 559–569, Elsievier.

[5] Detecting evolutionary financial statement fraud Wei Zhou , Gaurav Kapoor , Decision Support Systems 50 (2011) 570–575, Elsievier.

[6] A computational model for financial reporting fraud detection Fletcher H. Glancy , Surya B. Yadav, Decision Support Systems 50 (2011) 595–601, Elsievier.

[7] An Investigation of the Negative Selection Algorithm for Fault Detection in Refrigeration Systems, Dan W Taylor1,2 and David W Corne1, ICARIS 2003, LNCS 2787, pp. 34-45 2003, Springer.

[8] Stephanie Forrest, Alan S Perelson, Lawrence Allen and Rajesh Cherukuri "Self-Nonself Discrimination in a Computer", *Proceedings of the IEEE Symposium on Research in Security and Privacy*, IEEE Press *(1994)*

[9] Revisiting Negative Selection Algorithms, Zhou Ji , AutoZone, Inc., Memphis, TN 38103, USA.Dipankar Dasgupta dasgupta@memphis.eduDepartment of Computer Science, The University of Memphis, Memphis, TN 38152,USA