Applied Soft Computing 24 (2014) 40–49

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A novel model for credit card fraud detection using Artificial Immune Systems

Neda Soltani Halvaiee*, Mohammad Kazem Akbari

Amirkabir University of Technology, 424 Hafez Street, Tehran, Iran

ARTICLE INFO

Article history:
Received 3 February 2013
Received in revised form 4 December 2013
Accepted 24 June 2014
Available online 15 July 2014

Keywords:
Credit card fraud detection
Artificial Immune Systems
Artificial Immune Recognition System
Memory cell
Cloud computing
MapReduce

ABSTRACT

The amount of online transactions is growing these days to a large number. A big portion of these transactions contains credit card transactions. The growth of online fraud, on the other hand, is notable, which is generally a result of ease of access to edge technology for everyone. There has been research done on many models and methods for credit card fraud prevention and detection. Artificial Immune Systems is one of them. However, organizations need accuracy along with speed in the fraud detection systems, which is not completely gained yet. In this paper we address credit card fraud detection using Artificial Immune Systems (AIS), and introduce a new model called AIS-based Fraud Detection Model (AFDM). We will use an immune system inspired algorithm (AIRS) and improve it for fraud detection. We increase the accuracy up to 25%, reduce the cost up to 85%, and decrease system response time up to 40% compared to the base algorithm.

1. Introduction

Credit card fraud is an important issue and has considerable cost for banks and card issuer companies. Financial organizations try to prevent account misuse using different security solutions. The more complex the security solutions are, the more sophisticated fraudsters get i.e. fraudsters change their methods over time. Therefore it is crucial to improve fraud detection methods along with security modules which try to prevent fraud. Fraud detection has become a crucial activity in order to decrease the impact of fraudulent transactions on service delivery, costs, and reputation of the company. There are plenty of methods used for fraud detection each of which tries to retain maximum quality of service while keeping false alarm rate at minimum. Fraud is cost and detecting it before the transaction is registered will reduce this cost significantly, which needs a very accurate system with quite few false alarms. Edge and Falcone Sampaio [1] state that while implementation of proactive methods increases the potential for early fraud alerting, real-time processing significantly reduces the available time window within which computational analysis should be performed and an accurate decision should be made in response to newly arriving transactions. The quicker a fraud detection system responds, the better. Fraud detection systems are trained using older transactions in order to decide about new ones. This training phase is time-consuming which can be parallelized in most cases. In order to reduce computation time one can reduce the number of previous transactions processed by minimizing the time window, use less complicated methods, and etc. each of which might result in reduction in accuracy, which means more missed fraud cases and more false alarms. Accordingly, a powerful tool is needed on which the fraud detection system could run and process transactions in minimum time This paper suggests using cloud computing i.e. implementing fraud detection system on a cloud-base file system, namely Hadoop, which makes data parallelization possible in large datasets.

Different methods have been used for fraud detection including Bayesian algorithm [2], Neural network [3], Markov model [4], account signature [1], Artificial Immune Systems [5–8]. AIS is based on human immune system and is similar to fraud detection system in many aspects: 1 – Both of them pursue the same goal of separating normal records from unauthorized ones. 2 – In both cases the number of normal records is much more than unauthorized ones. 3 – In both cases unauthorized records are similar to those of normal. In human body viruses and non-self cells carry protein and masquerade self cells. Fraudsters also try to have similar behavior to card owner's behavior. 4 – Both systems have to detect and learn new methods of misuse. Human body could face new types of non-self cells any time and it has to not only detect the new types, but also remember them so that they can be detected later. Similarly, a typical fraud detection system should be able to detect any type

* Corresponding author at: No. 21, Mohajeran Avenue, Imam Street, Azarshahr, East Azarbaijan, Iran. Tel.: +98 9143053404.
E-mail addresses: neda.soltani@aut.ac.ir, neda.soltany@gmail.com (N. Soltani Halvaiee), akbarif@aut.ac.ir (M.K. Akbari).

http://dx.doi.org/10.1016/j.asoc.2014.06.042
1568-4946/

of fraud even if it had not happened before. Also the system should learn it for future cases.

AIS addresses detecting non-self cells, imitating the functions of human body which occurs during generating detector cells, detecting non-self cells, and cleaning the body from non-self while learning its pattern. Detector cells, namely lymphocytes, are self-tolerant which means they are not stimulated by self cells but by non-self cells. Immune system can learn new patterns of non-self cells that it has not come across before. Once a detector is stimulated by a non-self, the system keeps the detector as a memory cell. Therefore, if that particular non-self cell enters body later, it can be detected again. This makes AIS adaptable to its environment. Immune system starts training with no information about non-self cells. This means it is trained using only self cells.

In this paper we will use an AIS-based method for credit card fraud detection and introduce AFDM. We will improve a previously introduced algorithm [9] in various aspects to get higher precision. We will also propose a new implementation model for the method in order to reduce training time. The results are compared to a similar work [6] which has improved AIS system parameters. We use the same parameters as well as the dataset used in [6]. The remainder of this paper is structured as follows: Section 2 presents the background information. First AIS is described followed by Artificial Immune Recognition System – the algorithm which is used in this paper. Then, after a brief introduction about Cloud Computing, Hadoop file system and MapReduce API are described. Section 3 is about related work in credit card fraud detection field focusing on using AIS for fraud detection. Section 4 describes AFDM, the methodology, the improvements on AIRS, and the implementation model. Section 5 includes the results of the tests. Finally, Section 6 discusses future research directions.

2. Background

2.1. AIS

AIS simulates human body immune system functionality. Human body detects non-self cells, which might be viruses, pathogens, germs, etc., by creating detector cells named lymphocytes. As this functionality is similar to what a typical fraud detection system does, AIS is used for fraud detection in some researches. AIS detects non-self cells using two basic functions in human body which generate and mature lymphocytes: Negative Selection and Clonal Selection. Detector cells (lymphocytes) are generated through random composition of protein patterns. Then they will be able to prevent any potential threat by covering many protein patterns randomly. In order to have self-tolerant detectors which do not react to self cells, the system declines those which do. It means that any randomly generated detector which detects a self cell dies immediately. Right after generation, detectors are presented to self cells and only those which do not react to self cells survive. This process is called Negative Selection. After this process the detectors enter the system and in their short life-time they are expected to face any potential non-self cell and detect it. If any detector detects a non-self cell, it can live longer in order to make body vaccinated against that non-self; the process is called Clonal Selection. When a detector comes across a non-self cell and detects it, the detector is cloned through mutation. One of the clones having the highest affinity with the non-self cell is selected as memory cell and lives longer in human body. If that specific type of non-self cell enters human body again, the system will detect it using memory cells.

In this paper we will make improvements on an AIS-based algorithm named Artificial Immune Recognition System. Watkins et al. [9] first introduced AIRS. It is a classification algorithm which uses Clonal Selection. In [10] the authors state following as some features of AIRS:

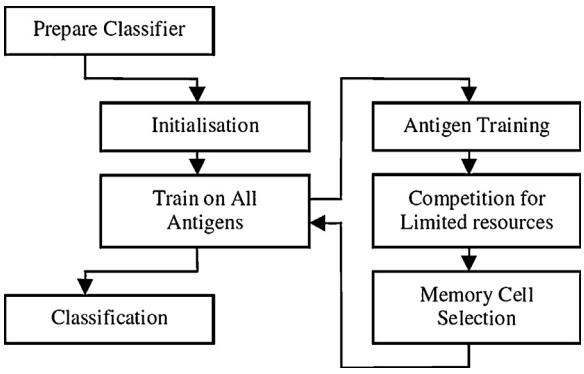

Fig. 1. Lifecycle overview of the AIRS algorithm [9].

- *Self-regulation*: AIRS does not require the user to select architecture, instead the adaptive process discovers or learns an appropriate architecture during training.
- *Performance*: Evaluation of the technique in some researches [6,10] show that AIRS is a competitive classification system.
- *Generalization*: Unlike techniques such as K-Nearest Neighbor that use the entire training dataset for classification, AIRS performs generalization via data reduction. This means that the resulting classifier produced by the algorithm represents the training data with a reduced or minimum number of exemplars. It is typical for AIRS to produce classifiers with half the number of training instances.
- *Parameter adjustment*: The algorithm has a number of parameters that allows tuning of the technique to a specific problem, with the intent of achieving improved results.

In AIRS both self/non-self cells and detector cells are represented as feature vectors. In order to reduce redundancy, ARB (Artificial Recognition Ball) is used which is representative of similar memory cells. ARBs are generated using a random mutation process with a certain probability and then ARBs that comply with the stop condition are selected as memory cells. Fig. 1 shows the flowchart of the algorithm. In the first step the classifier is prepared by normalizing training set records. Then the parameters of the algorithm are initialized and affinity threshold is calculated using the distance between any two records in training set. After this part some records are chosen randomly as initial ARBs. This part of the algorithm is time-consuming. The rest of the algorithm processes each single record. In this part, memory cells which have been stimulated the most by the training record (antigen) and their class attributes are the same as training record are chosen to be added to memory cell pool. Stimulation is calculated using the distance function. Choosing the memory cells is done by competition between existing memory cells and their clones. In the end a memory cell is chosen and added to memory cell pool and the training on a specific training record is done. The algorithm continues training on the rest of training records in the same way. This part is time-consuming, too. The last part of algorithm is Classification which starts when the training is done on all records. In this part K-Nearest Neighbor is used as the classifier algorithm. Each record in the test set is presented to memory cells and the neighbors are chosen based on stimulation. Then the class of the antigen is decided based on the class of the majority of K memory cells in the neighborhood.

2.2. Cloud computing

Cloud computing offers features which can help fraud detection in some aspects. First, cloud computing offers computation power: cloud computing uses datacenters with vast resources which has almost no limitation in computing, memory, and storage. Cloud