Journal of Applied Sciences Research, 4(6): 707-715, 2008

Predicting Fraud in Mobile Phone Usage Using Artificial Neural Networks

Oluwagbemi O.O

Department of Computer & Information Sciences, College of Science & Technology,
Covenant University, Ota, Ogun State, Nigeria.

Abstract: Mobile phone usage involves the use of wireless communication devices that can be carried anywhere, as they require no physical connection to any external wires to work. However, mobile technology is not without its own problems. Fraud is prevalent in both fixed and mobile networks of all technologies. Frauds have plagued the telecommunication industries, financial institutions and other organizations for a long time. The aim of this research work and research publication is to apply 3 different neural network models (Fuzzy, Radial Basis and the Feedforward) to the prediction of fraud in real-life data of phone usage and also analyze and evaluate their performances with respect to their predicting capability. From the analysis and model predictability experiment carried out in this scientific research work, it was discovered that the fuzzy network model had the minimum error generated in its fraud predicting capability. Thus, its performance in terms of the error generated in this fraud prediction experiment showed that its NMSE (Normalized mean squared error) for the fraud predicted was 1.98264609. The mean absolute error (MAE = 15.00987244) for its fraud prediction was also the least; this showed that the fuzzy model fraud predictability was much better than the other two models.

Keywords: Mobile phone, wireless communication, fraud prediction, fuzzy, radial basis, feedforward

INTRODUCTION

Mobile Technology makes use of wireless communication devices that can be carried anywhere, as they require no physical connection to any external wires to work.

A new era of mobile multimedia applications and services has been brought about by the the rapid growth of Wireless Communication and access, together with the success of the internet.

However, mobile technology is not without its own problems. Fraud is prevalent in both fixed and mobile networks of all technologies. Fraud can be defined as criminal deception. It is the use of false representation to gain an unjust advantage. Johnson[8] defines fraud as any transmission of voice or data across a telecommunication network where the intent of the sender is to avoid or reduce legitimate call charges

Frauds have plagued telecommunication industries, financial institutions and other organizations for a long time[7]. Fraudsters are not only motivated by money, but also by the need for anonymity to mask other crimes or the challenge of beating a security system. Fraudsters are ingenious and determined. They frequently find a way to misuse services and compromise security.

As a result, fraud counter-measures is essential and there is need for corporate policy supported by a suitable implementation strategy not only in combating mobile technology crimes, but also in the prediction of fraud occurrence in mobile phone usage.

The power of neural network-based technology is a potent mechanism in combating the menace of fraud in mobile technology. Thus, the importance of neural networks cannot be overemphasized in dealing with fraud related issue in mobile technology. Neural networks have over the years employed different strategies in combating this problem.

Different neural network models have been developed for the development of efficient fraud predicting systems. These systems can predict fraud through pattern recognition techniques and other neural networks related techniques[1].

The problem of the fraudulent use of mobile phones especially in Nigeria and in some other African countries is not an uncommon thing in wireless communication. Superimposed fraud is a kind of fraud where fraudsters can take over a legitimate account. In such cases, the abnormal usage of the mobile phone is superimposed upon the normal usage of the legitimate customers. The outcome of this process is that, the account of the legitimate user becomes more debited than the normal amount consumed. The legitimate user

Corresponding Author: Oluwagbemi O.O, Department of Computer & Information Sciences, College of Science & Technology, Covenant University, Ota, Ogun State, Nigeria.
Email:gbemiseun@yahoo.com

becomes inhibited to the proper and efficient use of such mobile phones, the timings in the bill record of the legitimate user becomes irregular and the overall security of all legitimate subscribers in such communication industry is not guaranteed.

The aim of this research work and research publication is to apply different neural network models to the prediction of fraud in real-life data of phone usage and also analyze and evaluate their performances with respect to their predicting capability.

Recently in Nigeria, there has been a wide usage of mobile phones. Thus, the problem of superimposed fraud may arise. Presently, the laws governing mobile technology in Nigeria does not adequately cater for this particular kind of fraud problem and also, no existing and tangible work has been done, hence a justification for embarking on such a scientific research work as a solution to this problem.

One of the objectives of this scientific research work is to demonstrate how different neural network models can be used in predicting fraud in mobile phone usage. It further includes determining the best model in terms of the least errors generated in the graphical analysis depicting the actual and predicted fraud.

MATERIALS AND METHODS

Relevant Materials and Previously Work Done: Fraud Detection In Communications Networks Using Neural And Probabilistic Methods: In this review, three methods to detect fraud were presented. Firstly, a feed-forward neural network based on supervised learning was used to learn a discriminative function to classify subscribers using summary statistics. Secondly, a Gaussian mixture model was used to model the probability density of subscribers' past behavior so that the probability of current behavior can be calculated to detect any abnormalities from the past behavior.

Lastly, the Bayesian network was used to describe the statistics of a particular user and the statistics of different fraud scenarios. The Bayesian network was now used to infer the probability of fraud given the subscribers' behavior. The data features were derived from toll tickets. The experiments showed that the methods detected over 85 % of the fraudsters in the research carried out by Michiaki Taniguchi and others in [1].

The main difference between this reviewed work and this research work is the probabilistic methods employed. The reviewed work is related to this research work because it employed the Feed forward Neural Network based on supervised learning to classify the phone subscribers and the Bayesian Neural Network to infer the probability of fraud, hence detecting fraudulent practices in the use of mobile phones.

Distributed Management in the Security Area for Cloned Mobile Phones: This work presented the development of a distributed application in the security management area for telecommunication networks. The application consists of a system that intends to avoid the use of cloned telephones. The main focus of this work was the classification of the telephone users into seven classes according to their usage logs. Such logs contained three relevant characteristics for every call made by the user. The classification made use of pattern recognition techniques. The users include: those that made few local calls, those with many local calls, those with few long distance calls, those with many long-distance calls, those with few short international calls, those with few long international calls, those with many long international calls.

It was possible to identify more easily if a call does not correspond to the patterns of a specific user, and thus identify, whether a fraudulent person made the call.

In order to conduct the classification from the existing data base, an artificial neural network was used, built from a Radial Base Function (Gaussian), known in the literature as RBF with the use of a clustering algorithm (k-means), which proved to be very efficient. Mirela Sechi Moretti Annoni Notare carried out this work in[2].

The relevance of this revised work to ours is about the use of the Radial Basis Function Neural Network model, as a means of detecting fraudulent practices in the use of mobile phone usage.

The work contributed towards the reduction of losses or damages, for clients as well as for telecommunication carriers, through the implementation of an anti-fraud system, which avoided the cloning of mobile phones. Beyond this, the work also employed a classification algorithm of high reliability. The method used for the classification of users, which included the K-means, P-Nearest Neighbour and Gauss algorithms and the purelin function, proved to be efficient and reliable with the use of the Mat Lab software.

Robust and Adaptive Travel Time Prediction With Neural Networks: In this work carried out by S.P. Hoogendoorn in 2000, an overview of the state -of-the art in travel time prediction using neural networks was presented[3].

However, prediction of travel times based on past and current traffic information is not straightforward due to the high complexity and ill -predictability of the traffic process, incorrect or missing observations, and

different data sources. Given the properties of the travel time prediction problem, neural networks were used as a tool for travel time prediction, as shown by a number of past studies including this reviewed work.In this work, two methods for predicting traffic conditions and times were presented. The indirect method dealt with the prediction of traffic conditions, while the direct method dealt with the prediction of travel times.

This work presented an overview of applications of ANN (Artificial Neural Networks) for travel time prediction. The applications showed very promising results.

A Comparison of Feed Forward Neural Network Architectures for Piano Music Transcription: The work carried out by Matija Marolt in 1999 showed the application of the feed forward neural networks in recognizing piano chords and polyphonic piano music transcription[4]. The work presented results obtained by using several feed forward neural network architectures for transcription, namely multilayer perceptrons, RBF networks, support vector machines and time-delay networks.

Pattern Recognition:

Radial Basis Function Network: The International School on Gas Sensors originally carried out this work in conjunction with the 3rd European School of the Nose Network in 2001. The main work done here, centered on gas and odour concentration prediction, using the radial basis network[5].

In gas and odour analysis, it was necessary to perform not only classification of substances but also quantification of the concentration level of individual components.

This was an important problem in environmental pollution monitoring and in cosmetics, food and industries.

The Radial Basis Function Network was made use of and the results obtained showed that the research work was successful.

Types of Fraud:

Fraud can be classified into two categories namely:

- Subscription fraud
- Superimposed fraud

- Subscription fraud: In this kind of fraud, fraudsters obtain a phone account without having any intention to pay the bill. In such cases, abnormal usage occurs throughout the active period of the account. Such account is usually used for call selling or intensive self-usage. Also, into this category falls the case of bad debt, where customers who do not necessarily have fraudulent intentions but never pay a single bill.

- Superimposed fraud: Here, fraudsters “take over” a legitimate account. In such a case, the abnormal usage is superimposed upon the normal usage of the legitimate customers. Examples of such include cellular cloning, call card theft, and cellular handset theft. Usage volume (total number, duration or rated value of calls over a certain period) is crucial in establishing a fraud case.

The focus of this scientific research work is to analyze and evaluate the performances of three different neural network models in predicting fraud occurrence in mobile phone usage. The kind of fraud being treated in this research work is that of the superimposed fraud.

Methodology: The methodology employed in this research work included, data collection by survey from the telecommunication industry ,data entry, data training and data testing, using three different neural network models all embedded in a Neural Network Software called NeuroSolutions..

Data collection: (description of data and significance of choice of data)

The data used for the purpose of this experiment was obtained by survey from the telecommunication industry. The significance of the choice of this data can bee seen in the important features it contains which helped to facilitate the discovery and detection of fraud occurrence in its usage patterns.

Sample data from the telecommunication industry

Time Of Call	Duration of Call (Mins)	Destination Called	Destination No	Rate
8:23	1	China	00986532579033	99.00
8:40	5	South Africa	00927117880335	67.50
8:43	5	United Kingdom	009442087408050	99.00
9:01	1	Sierria Leone	009232222241039	67.50
9:14	2	United Kingdom	009447950364015	99.00

9:32	1	Italy	009393339145272	99.00
9:46	10	United Kingdom	009442083100189	99.00
9:47	5	United Kingdom	009442083100189	99.00
10:02	2	Benin	009229941278	67.50
10.08	3	Canada	00913182471323	67.50
10:30	4	Australia	0091297711764	99.00
11:09	2	Togo	0092289493998	99.00
11:13	5	United Kingdom	009442073576822	99.00
11:27	7	Sierra Leone	0092322222241350	67.50
11:35	9	Benin	009229335538	67.50
11:46	5	USA	00912122819185	99.00
11:52	1	Canada	00912142325592	99.00
11:53	3	USA	0092149046400	99.00
12:02	1	United Kingdom	009442087650981	99.00
12:26	1	United Kingdom	009442085338257	99.00

Data Preprocessing: (scaling techniques) The data collected by survey was scaled for it to properly fit into the neural network software used for this research work. The major factors discussed in the previous section was applied.

Identification of Factors That Can Be Useful in Fraud Detection: Certain factors could be very useful in determining and predicting the occurrence of fraud patterns in a mobile phone usage data.

These Factors Are:

- The call collision detection factor which identified overlapping calls.
- Irregularity or inconsistency in phone no-location relation
- The call duration factor, which monitored individual and aggregate calls for specific conditions such as calls which are unusually long.
- The content factor, which helps to monitor and note the occurrence of fraudulent context of words spoken during phone conversation.

All these factors helped in establishing fraud patterns in phone usage data. Based on these factors, the same data sample was fed into about nine different neural network models in order to analyze the predicting capability of each network model

Preprocessed data

Time Of Call	Duration of Call (Mins)	Destination Called	Destination No	Fraud Occurrence Value(FOV)
8:23	1	China	00986532579033	115
8:40	5	South Africa	00927117880335	230
8:43	5	United Kingdom	009442087408050	115
9:01	1	Sierria Leone	009232222241039	230
9:14	2	United Kingdom	009447950364015	100
9:32	1	Italy	009393339145272	100
9:46	10	United Kingdom	009442083100189	230
9:47	5	United Kingdom	009442083100189	220
10:02	2	Benin	009229941278	110

10.08	3	Canada	00913182471323	120
10:30	4	Australia	0091297711764	130
11:09	2	Togo	0092289493998	125
11:13	5	United Kingdom	009442073576822	220
11:27	7	Sierra Leone	0092322222241350	225
11:35	9	Benin	009229335538	220
11:46	5	USA	00912122819185	120
11:52	1	Canada	00912142325592	125
11:53	3	USA	0092149046400	225
12:02	1	United Kingdom	009442087650981	230
12:26	1	United Kingdom	009442085338257	228

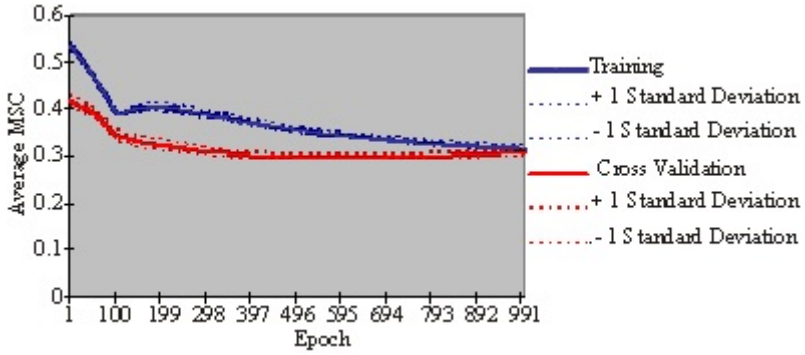
Regions of fraud occurrence take values btw 200-230 at random
Regions of Non-fraud occurrence take values btw 100-130 at random

Implementation: This research work was implemented by using a Neural network software called NeuroSolutions. The phone data was keyed into three different neural network models.

These data were fed into three different neural network models for prediction purpose.

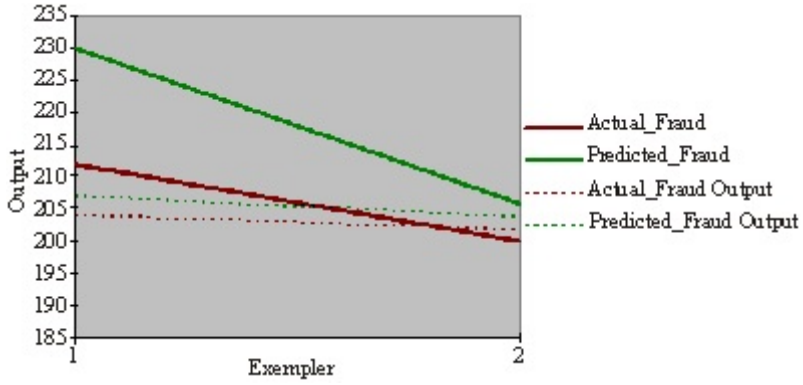
Simulation Results & Models Performances: Report on the experimental results obtained and the display of relevant graphs are shown in the appendix.

Simulation Results & Analysis of Models Performances:
Fuzzy Network Model Training & Performance Evaluation:



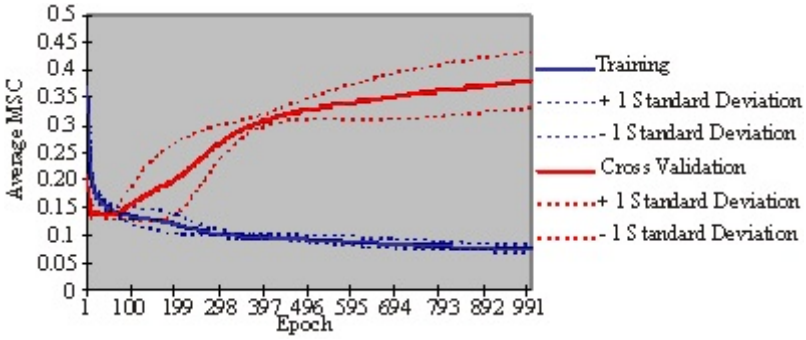
All Runs	Training Minimum	Training Standard Deviation	Cross Validation Minimum	Cross Validation Standard Deviation
Average of Minimum MSEs	0.316919863	0.005108523	0.29707703	0.005664798
Average of Final MSEs	0.316919863	0.005108523	0.308057845	0.006062931

Fuzzy Network Model Fraud Prediction Analysis:



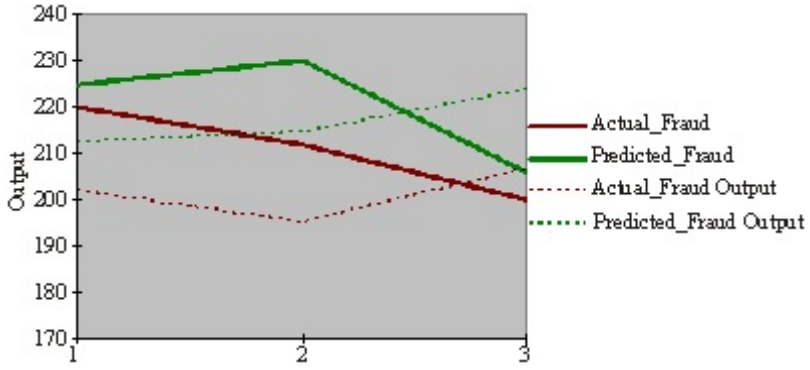
Performance	Actual_Fraud	Predicted_Fraud
NMSE	0.891501691	1.98264609
MAE	4.727737427	15.00987244
Min Abs Error	1.606430054	2.180221558
Max Abs Error	7.8490448	23.11178589
R	1	1

Feed Forward Network Model Training & Performance Evaluation:



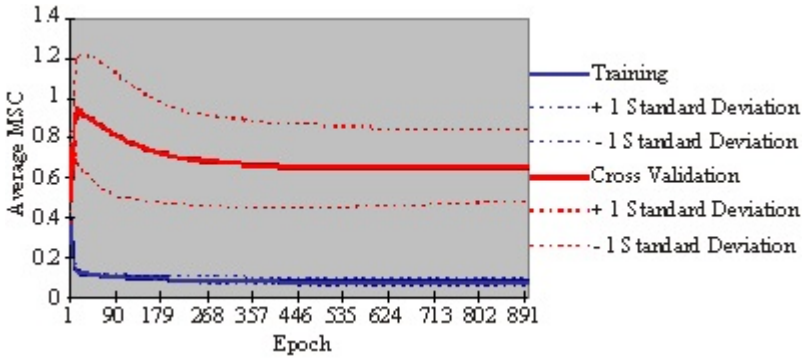
All Runs	Training Minimum	Training Standard Deviation	Cross Validation Minimum	Cross Validation Standard Deviation
Average of Minimum MSEs	0.073837444	0.005849992	0.130082786	0.001430886
Average of Final MSEs	0.073837444	0.005849992	0.381579846	0.051584337

FeedForward Network Model Fraud Prediction Analysis:



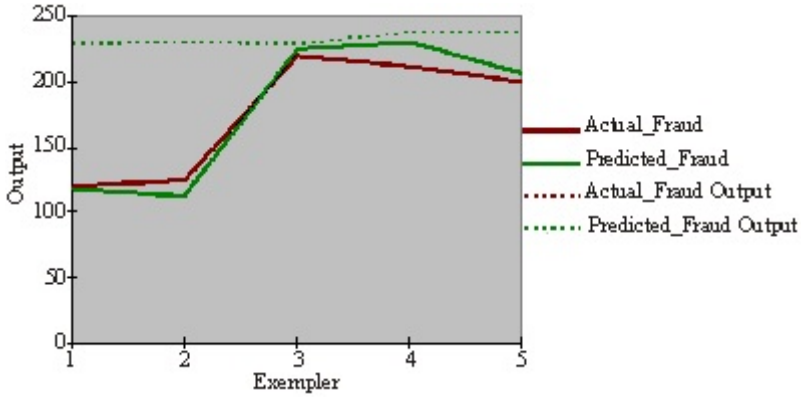
Performance	Actual_Fraud	Predicted_Fraud
NMSE	3.226233252	2.887962542
MAE	13.87532552	19.82550218
Min Abs Error	6.791488647	12.27384949
Max Abs Error	18.12606812	17.69332886
r	-0.527090207	-0.943488246

Radial Basis Function Model Training & Performance Evaluation:



All Runs	Training Minimum	Training Standard Deviation	Cross Validation Minimum	Cross Validation Standard Deviation
Average of Minimum MSEs	0.071652666	0.018137008	0.456480503	0.111232817
Average of Final MSEs	0.071652666	0.018137008	0.657245219	0.179131821

Radial Basis Function Model Fraud Prediction Analysis:



Performance	Actual_Fraud	Predicted_Fraud
NMSE	2.623272336	2.364837249
MAE	57.12963562	65.42110535
Min Abs Error	9.173934937	4.041503906
Max Abs Error	109.0605469	116.624176
R	0.541450345	0.597647241

Conclusion: From the analysis and model predictability experiment carried out in this scientific research work, it was discovered that the fuzzy network model had the minimum error generated in its fraud predicting capability. For the Fuzzy network model, the following was generated:

NMSE	0.891501691	1.98264609
MAE	4.727737427	15.00987244
Min Abs Error	1.606430054	2.180221558
Max Abs Error	7.8490448	23.11178589
R	1	1

Thus, its performance in terms of the error generated in this fraud prediction experiment showed that its NMSE (Normalized mean squared error) for the fraud predicted was 1.98264609.

The mean absolute error (MAE = 15.00987244) for its fraud prediction was also the least; this showed that the fuzzy model fraud predictability was much better than the other two models.

The minimum absolute error (Min abs Error= 2.180221558) and the maximum absolute error (Max abs Error= 23.11178589) was also the least in fraud prediction which showed that the fuzzy model was more effective and efficient in its fraud prediction capability than the other two models (Feed forward model, Radial Basis model).Thus, the fuzzy network model outperformed these other models in terms of the errors generated in their fraud predicting capability.

For Radial Basis Function,the following values was produced:

NMSE	2.623272336	2.364837249
MAE	57.12963562	65.42110535
Min Abs Error	9.173934937	4.041503906
Max Abs Error	109.0605469	116.624176
R	0.541450345	0.597647241

For the Feedforward model, the following was produced:

NMSE	3.226233252	2.887962542
MAE	13.87532552	19.82550218
Min Abs Error	6.791488647	12.27384949
Max Abs Error	18.12606812	17.69332886
r	-0.527090207	-0.943488246

REFERENCE

1. Michiaki, T., H. Michael, H. Jaakko and T. Volker, 1998. Fraud Detection in Communications Networks using Neural and Probabilistic Methods, Siemens AG, Corporate Technology Department Information and Communications D-81730 Munich, Germany.
2. Mirela Sechi Moretti A.N., *et. al* 1998. Distributed Management in the Security Area For Cloned Mobile Phones, Federal University of Santa Catarina, Brazil.
3. Hoogendoorn, S.P and. J.W.C van Lintin Ir, 2000. Robust and Adaptive Travel Time Prediction with Neural Networks, TRAIL Research School, Delft University of Technology, August.
4. Matija, M., 1999. A comparison of feed forward neural network architectures for piano music transcription, Faculty of Computer and Information Science University of Ljubljana, Slovenia.
5. S. Caesarea, T.L., 2001. Pattern Recognition-Radial Basis Function Networks, International School On Gas Sensors in conjunction with the 3rd European School of the Nose Network, June 2001.
6. Richard, J. Bolton and J. David Hand, January 2002. Statistical Fraud Detection: A Review.
7. Jia Wu and Jongwoo Park, 2001. Intelligent Agents and Fraud Detection.
8. Johnson, M., 1996. Causes and Effect of Telecoms Fraud, Telecommunication (Interrnational Edition), 30(12): 80-84.
9. Adrian, G. Bors Introduction Of the RBF (Networks) 2001. Dept. Of Computer Science, University of York, SDD, U.K.
10. Haykins, S., 1994 Neural Networks; A Comprehensive Foundation, Upper Saddle River, NJ: Prentice Hall.
11. Park, J. and J.W. Sandberg, 1991. “Universal Approximation Using Radial Basis Function Network” Neural Computation, 3: 246-257.
12. Poggio, T. and F. Girosi, 1990. ”Networks for approximation and learning” Proc. IEEE, 78(9): 1481-1497.