

Anomaly Detection Using Lion Optimization Algorithm in Mobile Network Data

^[1] T S Prabhakar, ^[2] Dr. Veena M N, ^[3] Likhitha K Prakash, ^[4] Namratha S, ^[5] Kalpitha K

^[1] Assistant Professor, Department of Information Science and Engineering.

^[2] Professor, PET Research Center.

^[3] ^[4] ^[5] Student, Department of Information Science and Engineering.

Corresponding Author Email: ^[1] sadiya.pa@gmail.com, ^[2] smanjupr@gmail.com

Abstract— Mobile network data contains the information of users. Metaheuristic algorithms are very popular for solving many engineering problems. In the proposed work CDR dataset, in which the smsin data column is utilized to find out the anomalies. Lion optimization algorithm, which is a metaheuristic population based is used for detecting the anomalies. In which the behaviours like roaming, hunting etc...Are used for optimization, where the proposed work has achieved the accuracy of 95%. Here SVM and Random forest algorithms are used for classifying the detected anomalies from optimization algorithm. The detected anomalies are visualized through graphs for better analysis.

Keywords: Anomaly, Metaheuristic

I. INTRODUCTION

A mobile network is an end to end communication network which is widely wireless [2]. As the economy and the lifestyle is increasing day by day. In recent years due to the development of 4G and 5G networks, along with the features and flexibility provided by these networks has gained more users mobile phone devices [5]. As the usage of mobile phone devices is also increasing at a high rate, this generates more data. This call detail record [CDR] data is one of the important and confidential information to the telecom industry, which is also used by researchers for various purposes [4]. The data generated is of event driven mobile phone data.

Anomalies are the behaviours which vary significantly from the majority of the data successful detection of anomalies can be used for various purposes like fraud detection, intrusion detection, crime investigation etc... [15] The presence of anomalies in any data becomes an issue, detection of these anomalies or unusual behaviour might be helpful for various purposes, such as security, recourse allocation for improving the network issues etc...So detection of these anomalies becomes more important.

In recent years metaheuristic algorithms are used for solving complex problems. Metaheuristic algorithms are bio/nature inspired algorithms which are mostly influenced by Darwin's evolution theory and are widely in use, these algorithms use genetic concepts[5].

II. LITERATURE REVIEW

Aryan Mokhtari [2] have utilized CDR dataset, and used three different models to detect anomalous data. In this approach a combination of methods such as GARCH,

K-means and Neural Network is used as a hybrid model and ARIMA model is used for time series forecasting. Xi Peng and others [4] experimented to improve the retrieval speed along with high compression ratio. Etienne Thuillier [5] used k-means clustering to detect the geographical location of users, by using large scale CDR dataset. [5] Have also found the movement of humans around the territory and have verified these results with three distinct National surveys. Kashif Sultan and others [8] have also used k-means clustering to find out the anomalies and have used ARIMA model for time series forecasting for future activity prediction of mobile users to predict future traffic in cellular networks by using CDR data. Tamer F. Ghanem [15] implemented a system for Intrusion detection by the method of anomaly detection using metaheuristic methods due to their high problem solving capability. [15] Have also compared the results with other algorithms such as Naive Bayes, Bayes Network, Bayesian Logistic Regression, Multilayer feedback neural network, Radial basis function network and Decision tree. [1] Has done a survey on one such metaheuristic algorithm that is Particle Swarm Optimization Algorithm [PSO] This algorithm is based on the swarms behaviour. This research provides a complete overview of the particle swarm optimization along with its variants, mechanism and the other features of this method. Nai Chun Chen and others [11] conducted a project using call detail records, to predict the next places visited by the tourists in the city of Andorra, Europe which was supported by the government. [11] Concluded that LSTM gives the best prediction of 97%. [3] have done outlier detection, where limited number of outlier examples are taken and by utilizing optimization approach is utilized for detecting anomalies. Data mining analyses on CDR data is done by using knowledge discovering algorithm LD-ABCD, which solves

many challenges [12]. [10] Has provided a review on important metaheuristic algorithms which uses new ideas along with various applications.

III. METHODOLOGY

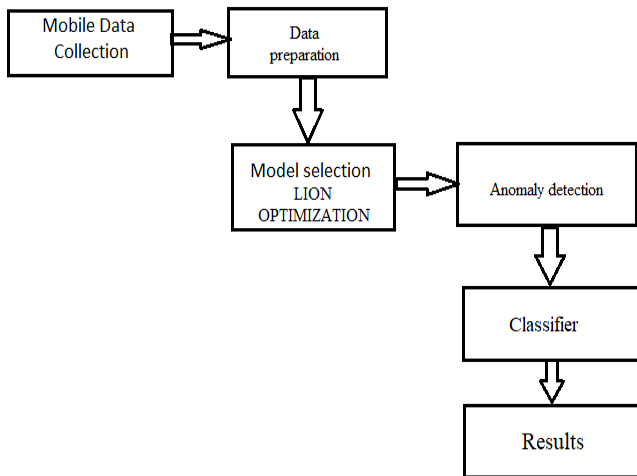


Figure 1: Methodology of the proposed work

Data Collection:

In the proposed work call detail record [CDR] dataset is been collected from the Telecom department of Milano city, Italy. This dataset contains the record of daily activity of users. This dataset consists of the data from the date 01-11-2013. It consists of about 10 lakh records. This dataset contains 8 attributes such as: Date and time, Cellid, countrycode, callin, callout, smsin, smsout and internet.

Dataset link: <http://dx.doi.org/10.7910/DVN/EGZHFV>

1	datetime	CellID	countrycod	smsin	smsout	callin	callout	internet
2	#####	1	0	0.3521			0.0273	
3	#####	1	33					0.0261
4	#####	1	39	1.7322	1.1047	0.5919	0.402	57.7729
5	#####	2	0	0.3581			0.0273	
6	#####	2	33					0.0274
7	#####	2	39	1.7334	1.088	0.602	0.4109	57.8875
8	#####	3	0	0.3644			0.0273	
9	#####	3	33					0.0287
10	#####	3	39	1.7348	1.0701	0.6128	0.4203	58.0095

Figure 2 : CDR dataset fields

Data preparation:

The CDR dataset contains many null values which are removed from this dataset. It contains the data columns like cellid and countrycode which are removed from the dataset as they are not used. The data contains of around 1048576 records in which only 139961 with 6 records are remained after data preparation. In this work 1000 records of smsin column are used for anomaly detection.

	datetime	smsin	smsout	callin	callout	internet
2	01-11-2013 00:00	1.7322	1.1047	0.5919	0.4020	57.7729
5	01-11-2013 00:00	1.7334	1.0880	0.6020	0.4109	57.8875
8	01-11-2013 00:00	1.7348	1.0701	0.6128	0.4203	58.0095
11	01-11-2013 00:00	1.7284	1.1533	0.5627	0.3763	57.4408
14	01-11-2013 00:00	1.5530	1.0439	0.5110	0.3503	52.1497
17	01-11-2013 00:00	1.7348	1.0701	0.6128	0.4203	58.0095
20	01-11-2013 00:00	1.7348	1.0701	0.6128	0.4203	58.0095
23	01-11-2013 00:00	1.7348	1.0701	0.6128	0.4203	58.0095
26	01-11-2013 00:00	1.7348	1.0701	0.6128	0.4203	58.0095
29	01-11-2013 00:00	0.9149	0.4990	0.4043	0.3287	33.5432

Figure 3: Dataset after preprocessing

Lion Optimization Algorithm:

Inspiration:

Lions reside socially, they usually reside in groups called “Prides” and some reside as “Nomads” [2]. Each pride consists of about five female lions, their cubs of both genders, one or more adult males and nomads who move irregularly in pairs or singularly [1]. These lions exhibit many characters like hunting, roaming around the territory, moving towards safe place etc... which are mathematically modelled to obtain the best solution.

Behaviours of Lions:

- **Hunting:** In every pride female lions look for hunt. These Lions use a different technique to catch the prey. They follow some pattern for hunting where the lions makes three different wings (centre, left, right) based on their fitness. The lions with the highest fitness form the centre wing.
- **Moving towards safe place:** In each pride some lions will go on hunting and the remaining lions will move towards the territory for encircling the prey.
- **Roaming:** Male lions in the pride roam around the territory which is considered as the current position of lion and the next visited position of the lion is considered as best visited position which is taken based on the fitness calculation.
- **Defence:** When male lions in the pride become mature, they fight with other male lions in the pride. The lion which loses will go out of the pride. In contrast nomad lions also attack the pride also.
- **Migration:** Lions exhibits the nature of switching lifestyles between nomads and prides.

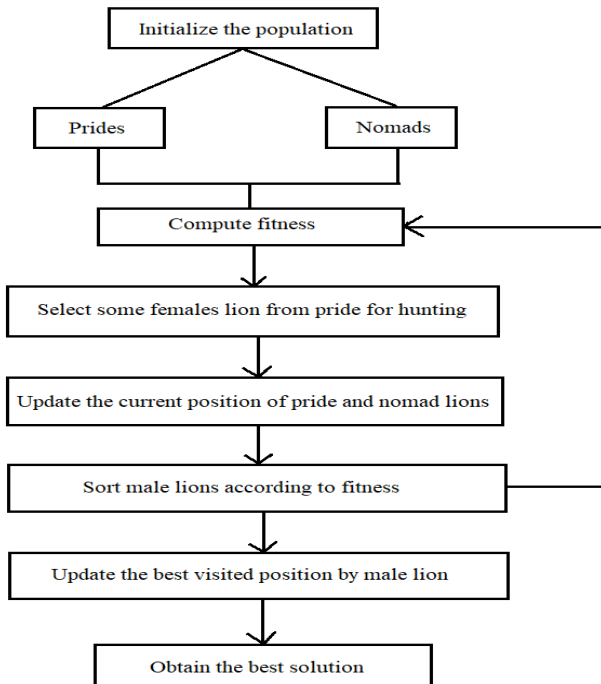


Figure 4 : Flow diagram of Lion optimization algorithm

IV. EXPERIMENTAL RESULTS

The proposed work detects the anomalies present in the CDR[Call Detail Record] dataset using lion optimization algorithm’s feature of roaming. The datapoints that are beyond the territory of the best visited positions of the lions are considered as anomalies. The SVM and Random forest classifiers are used for classifying the results of anomalies detected to form a matrix of 0’s and 1’s.

In the proposed work the optimizer has obtained the accuracy of about 95%. As it is an optimizer it takes the best solution every time over the iterations.

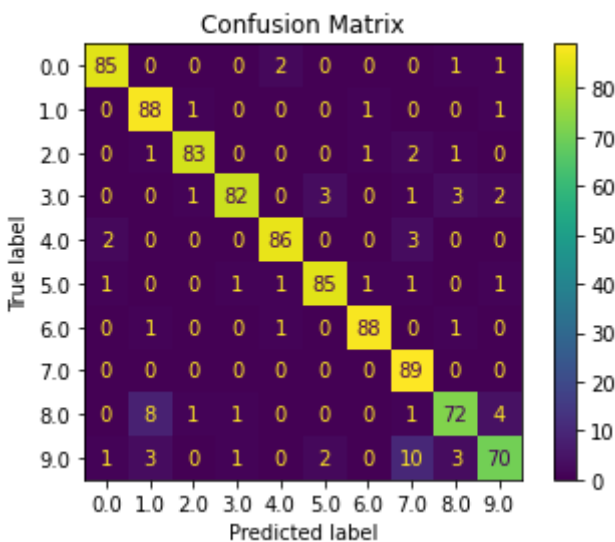


Figure 5: Confusion matrix of 95% accuracy

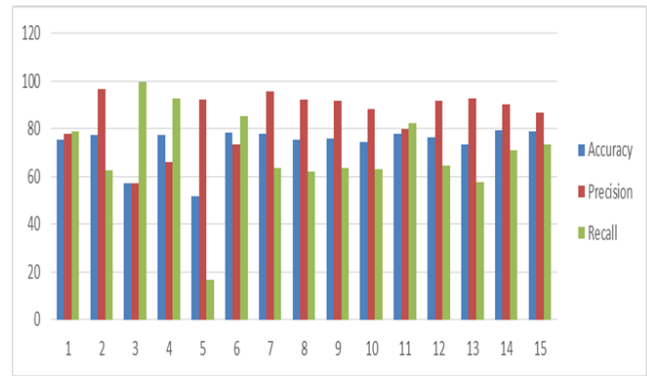


Figure 6 : Accuracy of the proposed work

These are the anomalies that are detected by using Lion Optimization Algorithm.

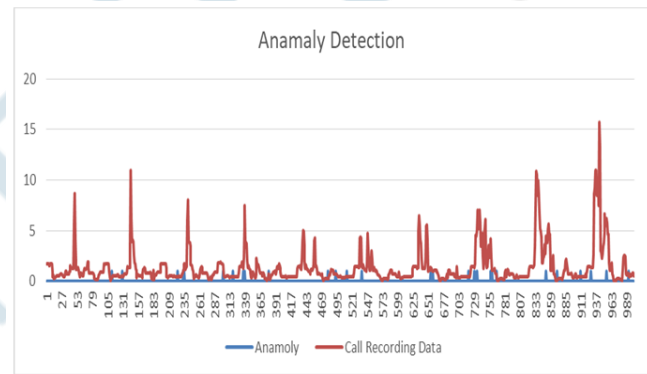


Figure 7 : Anomaly Detection in CDR data

V. CONCLUSION AND FUTURE ENHANCEMENT

In the proposed work CDR data is utilized to detect anomalies by using “Lion Optimization” algorithm where the accuracy of 95% is obtained for the best solution.

CDR anomaly detection can be helpful for improving the quality of service which can be achieved by providing efficient resource allocation and distribution. This anomaly detection can be used for various purposes like fraud detection, intrusion detection etc... And the anomalies can be detected for all the attributes present in the dataset and by utilizing the entire dataset. And other features of the Lion Optimization algorithm can be implemented to find out solutions for different problems.

REFERENCES

- [1] Ahamed G Gad [2021] “Particle Swarm Optimization Algorithm and its Applications: A Systematic Review” <https://link.springer.com/article/10.1007/s11831-021-09694-4>
- [2] Aryan Mokhtari, Leyla Sadighi, Behnam Bahrak, and Mojtaba Eshghie [1]2020.”HYBRID MODEL FOR ANOMALY DETECTION ON CALL DETAIL RECORDS BY TIME SERIES FORECASTING” Hybrid Model for Anomaly Detection on Call Detail Records by Time Series Forecasting | DeepAI

- [3] Niko Reunanen, Tomi Raty, Timo Lintonen [2020] "Automatic Optimization of Outlier detection ensembles using a limited number of outlier examples" <https://boi.org/10.1007/s41060-020-00222-4>
- [4] XI PENG, LIANG LIU, AND LEI ZHANG [2]2019 "Hive-Based Retrieval Optimization Scheme for Long-Term Storage of Massive Call Detail Records" A Hive-Based Retrieval Optimization Scheme for Long-Term Storage of Massive Call Detail Records | IEEE Journals & Magazine | IEEE Xplore
- [5] Etienne Thuilier, Laurent Moalic, Sid Ahmed Lamrous, and Alexandre Caminada [3]2019. "Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data" Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data (archives-ouvertes.fr)
- [6] Al-Thanoon NA, Qasim OS, Algamal ZY (2019) A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics. Chemom Intell Lab Syst 184:142–152
- [7] Alswaitti M, Albughdadi M, Isa NAM (2018) Density-based particle swarm optimization algorithm for data clustering. Expert Syst Appl 91:170–18
- [8] Kashif Sultan, Hazrat Ali, Zhongshan Zhang [4]2018 "Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks" <https://ieeexplore.ieee.org/document/8419685>
- [9] T. Andrysiak, L. Saganowski, M. Maszewski and A. Marchewka. Detection of Network Attacks Using Hybrid ARIMA-GARCH Model. In Proceedings of the Twelfth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, Jul 2018, pp.1-12.
- [10] Arockia Panimalar.S [2017] "Nature inspired Metaheuristic Algorithms" <https://www.irjet.net/archives/V4/i10/IRJET-V4i1054>
- [11] Nai Chun Chen, Wanqin Xie, Jenny Xie, Kent Larson and Roy E. Welsch [5]2017 "Comprehensive Predictions of Tourists' Next Visit Location Based on Call Detail Records using Machine Learning and Deep Learning methods" <https://ieeexplore.ieee.org/document/80293>
- [12] Filippo Maria Bianchi, Antonello Rizzi, Alireza Sadeghian and Corrado Moiso [2017] "Identifying user habits through data mining on call data records" https://www.researchgate.net/publication/303507139_Identifying_user_habits_through_data_mining_on_call_data_records
- [13] M. S. Parwez, D. B. Rawat and M. Garuba. Big Data Analytics for User-Activity Analysis and User-Anomaly Detection in Mobile Wireless Network. IEEE Transactions on Industrial Informatics, 2017, 13(4): 2058-2065.
- [14] Maziar Yazdani, Faribroz Jolai [2015]. "Lion Optimization Algorithm(LOA) : A nature inspired metaheuristic algorithm" <https://www.sciencedirect.com/science/article/pii/S2288430015000524>
- [15] Tamer F. Ghanem , Wail S. Elkilani , Hatem M. Abdul-kader [2015] "A Hybrid Approach for Efficient Anomaly Detection using Metaheuristic Methods" <http://dx.doi.org/10.1016/j.jare.2014.02.009>
- [16] E. Letouze, "Key Facts & Figures on Big Data and Call Detail Records," 2014, Accessed Jan 2017. [Online]. Available: <http://datapopalliance.org/key-facts-figures/>
-