International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 6th Mar 2018. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-03

Title: IDENTIFYING ANOMALIES IN OSN BY USING NHAD MECHANISM

Volume 08, Issue 03, Pages: 13–17.

Paper Authors

MR.K.HARIKRISHNA, G.DURGA MAHESH

Vignan's Lara Institute of Technology & Science

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IDENTIFYING ANOMALIES IN OSN BY USING NHAD MECHANISM

MR.K.HARIKRISHNA[1], G.DURGA MAHESH[2]

Assistant Professor[1], Department of M.C.A, Vignan's Lara Institute of Technology & Science

M.C.A Student[2],Department of M.C.A ,Vignan's Lara Institute of Technology & Science

Abstract:

Use of social network is the basic functionality of today's life. With the advent of more and more online social media, the information available and its utilization have come under the threat of several anomalies. Anomalies are the major cause of online frauds which allow information access by unauthorized users as well as information forging. One of the anomalies that act as a silent attacker is the horizontal anomaly. These are the anomalies caused by a user because of his/her variable behaviour towards different sources. Horizontal anomalies are difficult to detect and hazardous for any network. In this paper, a self-healing neuro-fuzzy approach (NHAD) is used for the detection, recovery, and removal of horizontal anomalies efficiently and accurately. The proposed approach operates over the five paradigms, namely, missing links, reputation gain, significant difference, trust properties, and trust score. The proposed approach is evaluated with three datasets: DARPA'98 benchmark dataset, synthetic dataset, and real-time traffic. Results show that the accuracy of the proposed NHAD model for 10% to 30% anomalies in synthetic dataset ranges between 98.08% and 99.88%. The evaluation over DARPA'98 dataset demonstrates that the proposed approach is better than the existing solutions as it provides 99.97% detection rate for anomalous class. For real-time traffic, the proposed NHAD model operates with an average accuracy of 99.42% at 99.90% detection rate.

Introduction

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to

the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

Existing system:

The existing solutions can resolve the anomalies using the network activity rather than the users' approach towards a particular source. Evaluations performed on the basis of network activity can give inaccurate results as users' network activity can be intentional or unintentional, whereas the users' continuous interaction with a particular source can give more details about its behavior in online social networks. Solutions like COPRA and Bayesian anomaly detection are available for the detection of anomalies in online social networks. The Bayesian approach utilizes the Bayesian filtering mechanism to identify the anomalous node in the social network, whereas COPRA deals with the identification of the overlapping communities in the social networks. COPRA can be used to identify anomalies by determining the users in the non-overlapping communities. Although these approaches are effective, they are unable to provide recovery and eradicate mechanisms. Existing neuro-fuzzy approaches like Mobile Fuzzy Trust Inference, Modularity maximization and Hybrid Genetic Detection can also be extended for detecting different users in a given social network. At present, these approaches are only evaluated for identifying trust between two users and for community detection. On a broader version, these approaches can be integrated with anomaly detection mechanism and their existing communication classification can be used for detecting horizontal anomalies. But, this may increase the complexity of the overall system. Some other solutions include co-clustering based collective anomaly detection using network patterns, and self-learning intrusion detection systems that use Radial Basis Functions (RBF) neural network to resolve anomalies. Also, there

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

are many approaches that primarily focus on deploying Support Vector Machine (SVM) along with other ideologies to detect anomalous behavior. Some of these are anomaly detection with principal component analysis and SVM, autonomous labeling with SVM, and ensemble technique for anomaly detection which uses SVM in combination with the Extended Kalman Filter. Although, the performance results of these solutions over standard benchmarks suggest their efficiency, yet these do not contain appropriate features of online social networks which are required for the detection of horizontal anomalies

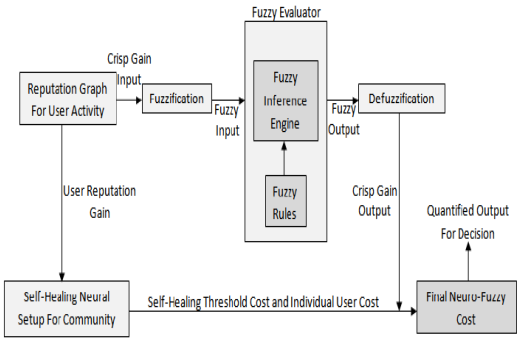
Proposed system

The proposed NHAD model aims at labeling a particular user in a community of an online social network to be an anomaly or not. NHAD uses the existing self-healing neural model for initializing the anomaly detected as a dummy neuron in the neural setup of the communities of an online social network. Then, this neural model heals using a fuzzy inference system with possibilities of recovering a user before completely eradicating it. The reputation gain of each user acts as a weight, and a healing cost is computed for each of the users. This healing cost is then used to find the final outcome for a node's activity; i.e. either an anomaly or a genuine user. For healing model application, the model is categorized as the neural setup shown in Fig. 3. The neural setup accounts for the m number of users in the jth community each treated as an input neuron with weight equivalent to their reputation gain. The hidden layer ("sources" in Fig. 3) is formed from the sources based on the user activity. The output of the neural model produces a threshold cost below which the user is treated as an anomaly. The final cost of a user is calculated after Defuzzification of the fuzzy set over Tp.

Advantages:

- A faster convergence approach despite the number of anomalies, fewer iterations to mark a user as an anomaly and smaller effect on the network activity. Further, the proposed NHAD model shows improvement in the convergence cost and the accuracy in the detection of a horizontal anomaly.
- Neuro-fuzzy solution for the identification of anomalies and system learning.
- Recovery after detection of horizontal anomalies.

Architecture:



The proposed NHAD model aims at labeling a particular user in a community of an online social network to be an anomaly or not. NHAD uses the existing self-healing neural model for initializing the anomaly detected as a dummy neuron in the neural setup of the communities of an online social network. Then, this neural model heals using a fuzzy inference system with possibilities

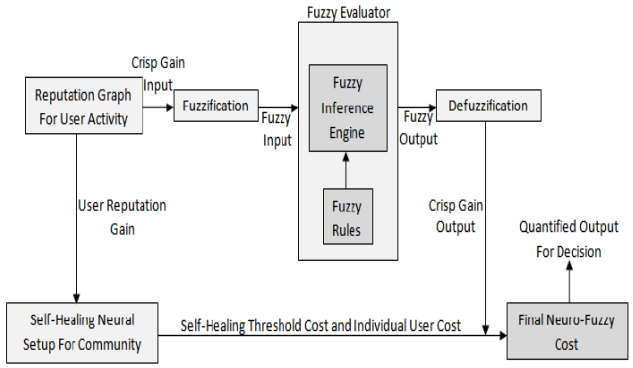
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

of recovering a user before completely eradicating it. The reputation gain of each user acts as a weight, and a healing cost is computed for each of the users. This healing cost is then used to find the final outcome for a node's activity; i.e. either an anomaly or a genuine user. For healing model application, the model is categorized as the neural setup shown in Fig. 3. The neural setup accounts for the m number of users in the jth community each treated as an input neuron with weight equivalent to their reputation gain. The hidden layer ("sources" in Fig. 3) is formed from the sources based on the user activity. The output of the neural model produces a threshold cost below which the user is treated as an anomaly. The final cost of a user is calculated after Defuzzification of the fuzzy set over Tp.

Advantages:

- A faster convergence approach despite the number of anomalies, fewer iterations to mark a user as an anomaly and smaller effect on the network activity. Further, the proposed NHAD model shows improvement in the convergence cost and the accuracy in the detection of a horizontal anomaly.
- Neuro-fuzzy solution for the identification of anomalies and system learning.
- Recovery after detection of horizontal anomalies.

Architecture:



Modules:

Reputation graph:

The proposed approach forms the reputation graph and then uses the fuzzy system to evaluate each user over the considered properties for their activities in a social network. Next, this reputation graph is used to find the self-healing cost of each user. Following this, the threshold healing cost, individual cost and crisp outcomes of each user are used to find the final neuro-fuzzy cost, which is used to decide whether a user is an anomaly or not.

Healing Cost and Neuro-Fuzzy Formations:

The first step in the proposed NHAD model is to map the defined set of properties to the neural network which operates by using a healing cost. The mapped network is then operated on the fuzzy inference rules to generate the fuzzy sets for the behaviour of each node, which is then evaluated to arrive at a decision of declaring a node as an anomaly or not.

Significant Difference:

It is based on the pattern of interaction between the two entities, and it helps in identification of a user as an anomaly. Significant difference controls the users'

reputation gain and its activity over the social media. The significant difference is much affected by the user activity over unverified sources. In this paper, the normalized controlling threshold deviation of a user in a community is fixed at a threshold of 0.5. This value is fixed considering that at the most a network can have 50% anomalies. Although in a real network, this value is very low, yet to prove the effectiveness of the proposed approach, a higher anomaly rate is chosen.

Reputation Gain:

Reputation gain Rg is computed over a graph G such that G=(Tp; Ts), where Tp denotes the set of trust properties that form the vertices of the graph, and Ts is the set of trust score assigned as weight to the edges connecting the vertices (Trust Properties) to a particular user

Conclusion:

The proposed approach was evaluated in three parts. The first evaluated the proposed NHAD model using a DARPA'98 dataset as used by most of the binary classification solutions, the second part evaluated it using synthetic dataset and the third part evaluated the proposed model over real-time traffic. The healing cost strategy of the proposed NHAD model allowed detection, recovery and removal decisions in fewer iterations, thus, making it an efficient scheme for the detection of horizontal anomalies in online social networks.

References

Oracle

PL/SQL Programming by Scott Urman

SQL complete reference by Livion

JAVA Technologies

JAVA Complete Reference

Java Script Programming by Yehuda Shiran

Mastering JAVA Security

JAVA2 Networking by Pistoria

JAVA Security by Scotl oaks

Head First EJB Sierra Bates

J2EE Professional by Shadab siddiqui

JAVA server pages by Larne Pekowsley

JAVA Server pages by Nick Todd

HTML

HTML Black Book by Holzner

JDBC

Java Database Programming with JDBC by Patel moss.

Software Engineering by Roger Pressman