International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Inductive and Transductive Transfer Learning for Zero-day Attack Detection

Nerella Sameera, Andhavarapu Bhanusri, M. Shashi

Abstract: ***Upon application of supervised machine learning techniques Intrusion Detection Systems (IDSs) are successful in detecting known attacks as they use predefined attack signatures. However, detecting zero-day attacks is challenged because of the scarcity of the labeled instances for zero-day attacks. Advanced research on IDS applies the concept of Transfer Learning (TL) to compensate the scarcity of labeled instances of zero-day attacks by making use of abundant labeled instances present in related domain(s). This paper explores the potential of Inductive and Transductive transfer learning for detecting zero-day attacks experimentally, where inductive TL deals with the presence of minimal labeled instances in the target domain and transductive TL deals with the complete absence of labeled instances in the target domain. The concept of domain adaptation with manifold alignment (DAMA) is applied in inductive TL where the variant of DAMA is proposed to handle transductive TL due to non-availability of labeled instances. NSL_KDD dataset is used for experimentation.***

Keywords *: Inductive transfer learning, Manifold alignment, Transductive transfer learning, Transfer Learning, Zero-day attack.*

I. INTRODUCTION

Existing signature-based approaches of Intrusion Detection Systems (IDSs) cannot detect zero-day attacks because of lack of sufficient labeled instances of zero-day attacks and anomaly-based approaches of IDS detect zero-day attacks but results in high False Positive Rates (FPR's) [1]. To detect zero-day attacks with high accuracies and low FPR's classifiers built on available labeled instances in one domain should be used to classify the zero-day attacks of another related domain. This can be possible with the Transfer Learning (TL) [2] approach. TL can be used to leverage the learned knowledge from one domain to the related domain. Inductive TL and transductive TL are the two important types of transfer learning [3]. Inductive TL requires the presence of minimal labeled instances in the target domain whereas the transductive TL can work with no labeled instances in the target domain. For transferring the knowledge among the domains which are usually having different feature spaces and different marginal probability distributions, both the domains should be transformed into the common latent space and this process is called manifold alignment. This paper extends manifold alignment called DAMA [4] to transform the multiple domains in to the common latent space at the same time. DAMA can be applicable if both the domains have the same label spaces. DAMA requires the construction of mapping functions to make this transformation which can be done by further constructing four matrices namely similarity, dissimilarity, source and target structural matrices. DAMA supports only inductive TL as DAMA requires the presence of minimal labeled instances in the target domain for constructing the similarity and dissimilarity matrices. This paper explores the potential of inductive and transductive TL methods for detecting zero-day attacks. For implementing transductive TL, authors in this paper constructs similarity and dissimilarity matrices by using cluster correspondence procedures due to non- availability of labeled instances. The rest of the paper is organized into 4 sections. Proposed inductive TL approach is presented in section II, proposed transductive TL approach is presented in section III, experimentation details and results are discussed in section IV and lastly conclusion is presented in section V.

II. PROPOSED INDUCTIVE TL APPROACH FOR ZERO-DAY ATTACK DETECTION

Authors have proposed a inductive TL approach for detecting zero-day attacks. The proposed approach extends DAMA [4] for solving the problem of domain unification where the domains are having with heterogenous feature spaces and different marginal probability distributions. Considering p domains with the corresponding data matrix X_P having the shape $q_p \times n_p$, p mapping functions are generated to transform p domains into the latent space. These mapping functions having the dimensionality $(q_1+ q_2+.....+q_p) \times b$ are the non-zero eigen vectors of the generalized eigen value decomposition given in eq.1

$$Z(\mu L+L_s)Z^T\varkappa = \lambda Z L_d Z^T\varkappa \quad (1)$$

where Z, L, L_s, L_d are the matrices further obtained from data matrices X_P, structural matrices W_P, similarity W_S, dissimilarity matrices W_d respectively.

Similarity Matrix:

Similarity matrix of the domains is constructed by matching the instances with the same label and separating the instances with different label. The similarity matrix W_S with dimensionality $(n_1+.....+n_p) \times (n_1+.....+n_p)$ is given below.

Manuscript published on 30 September 2019.
*Correspondence Author(s)
Nerella Sameera*, Research Scholar, Dept. of CS&SE, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, India. Email: sameerascholar@gmail.com
Andhavarapu Bhanusri, Mtech Scholar, Dept. of CS&SE, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, India. Email: abhanusri6@gmail.com@gmail.com
M. Shashi, Professor, Dept. of CS&SE, Andhra University College of Engineering (A), Andhra University, Visakhapatnam, India. Email: smogalla2000@yahoo.com

$$W_S = \begin{bmatrix} W_s^{1,1} & \dots & W_s^{1,p} \\ \dots & \dots & \dots \\ W_s^{p,1} & \dots & W_s^{p,p} \end{bmatrix}$$

Where the sub matrix $W_s^{a,b}$, the value is taken 1 for the corresponding instances with the same class label and for the remaining cases the value taken as 0.

Dissimilarity Matrix:

Dissimilarity matrix of the domains is constructed by separating the instances with the same label and matching the instances with different label. The dissimilarity matrix W_d with dimensionality $(n_1+\dots+n_p) \times (n_1+\dots+n_p)$ is given below.

$$W_d = \begin{bmatrix} W_d^{1,1} & \dots & W_d^{1,p} \\ \dots & \dots & \dots \\ W_d^{p,1} & \dots & W_d^{p,p} \end{bmatrix}$$

Where for the sub matrix $W_d^{a,b}$, the value is taken 1 for the corresponding instances with the different class label and for the remaining cases the value taken as 0.

Structural Matrix:

Structural matrix represents the intra similarity of the specific domain. The intra similarity distance is calculated from the formula given bellow.

$W_p(i, j) = e^{-||xi-xj||^2}$

Once all the domains are transformed into the common latent space, they can be compared. Hence the classifier constructed using all the labeled data (labeled data from both source and target domains) can be used to classify the unlabeled target instances.

III. PROPOSED TRANSDUCTIVE TL APPROACH FOR ZERO-DAY ATTACK DETECTION

Since most of the times zero-day attacks may don't have minimal labeled instances, authors have extended the work mentioned in section II to allow target domain with no labeled instances. Since labels are completely absent in the target domain, the authors have constructed similarity and dissimilarity matrices by following cluster correspondence procedures and structural matrix can be constructed by following the same procedure discussed above in section II.

Similarity Matrix:

All instances of the dataset are grouped into k clusters. A rank is assigned to each cluster based on its (centroid) distance from the global mean, such that the higher rank is assigned to the cluster that is having less distance from the global mean. The algorithm of this ranking process is given in table 1.

Once the ranking process is done, the same rank should be assigned to the instances of the cluster. Similarity matrix of the domains a and b is constructed by aligning the corresponding clusters from both the domains. If i^{th} and j^{th} instances of the domains having the same rank then the value $W_s^{a,b}(i,j)$ is taken as 1 or else if the rank of j lesser or greater by 1 to the rank of i, then the value $W_s^{a,b}(i,j)$ is taken as 0.5 otherwise the value is taken as 0.The pseudo-code for the construction of similarity sub matrix $W_s^{a,b}$ is given in table 2.

Dissimilarity Matrix:

Dissimilarity matrix W_d representing the non-correspondence among the domains and is obtained by complementing the similarity matrix. If i^{th} and j^{th} instances of the domains having the same rank then the value $W_d^{a,b}(i,j)$ is taken as 0 or else if the rank of j lesser or greater by 1 to the rank of i, then the value $W_d^{a,b}(i,j)$ is taken as 0.5 otherwise the value is taken as 1.

Table 1: Algorithm for ranking clusters

Input: Dataset $X_i \in X$. where $X=\{X_1,X_2,\dots,Xp\}$
Output: Dataset X_i with cluster ranks assigned to its instances.
Algorithm:
Step 1: Calculate global mean μ_{Xi} of X_i
Step 2: Form k clusters $\{C_1,C_2,,,,,C_k\} \in C_{Xi}$ using k-medoids clustering method.
Step 3: For each cluster $C_i \in C_{Xi,}$ repeat the step (a).
(a)Calculate distance d_i between centroid C_i and $\mu_{Xi.}$
Step 4: Sort the clusters in ascending order based on the distances $d_1, d_{2,\dots,}d_k$.
Step 5: Rank the clusters according to the sorting order from 1 to k.
Step 6: Assign the cluster rank to the cluster instances.

Table 2: Pseudo-code for generating similarity submatrix $W_s^{a,b}$

Input: Datasets $\{a, b\} \in X$. With cluster ranks assigned to instances.
Output: Similarity sub matrix $W_s^{a,b}$
Pseudo-code:
Steps:

```
1.  Initialize the matrix W_s^{a,b} with zeros.
2.  for each instance i of dataset a
3.      for each instance j of dataset b
4.          r1=rank of i
5.          r2=rank of j
6.          if r2 = = r1
7.           then W_s^{a,b}{i, j} =1
8.           else if (r2=r1+1 or r2=r1-1)
9.                Then W_s^{a,b}{i,j}=0.5
10.           end if
11.         end if
12.     end for
13. end for
```

IV. EXPERIMENTATION AND RESULTS

A. Dataset

Encoded NSL_KDD [5] dataset is used to implement the proposed approach. All attack instances of the data set belong to four attack groups namely DoS, Probe, R2L and U2R. For the purpose of experimentation, two modules namely DoS_module and R2L_module are prepared from the dataset after scaling. DoS_module contains only instances belongs to DoS attack group along with some normal instances equal sized with DoS instances and R2L_module contains only instances belongs to R2L attack group along with some normal instances equal sized with R2L instances.

B. Experimentation of Inductive TL approach

DoS_module is considered as source and R2L_module is considered as target. The proposed approach considers only 10% labeled instances of the target. So, 90% labels were removed for R2L_module. Fully labeled DoS_module and 10% labeled R2L_module are given as input to the proposed Inductive TL approach for detecting zero-day attacks. K-Nearest Neighbor (KNN) classifier is used to classify the zero-day attacks. The best K value is selected as 3 by experimenting on the validation data and the corresponding experimentation process is given in Fig.1. KNN is applied to classify unlabeled target instances with the selected best k value and achieve an accuracy of 88.24% and FPR of 11.31%. The results are shown in the Table 3.

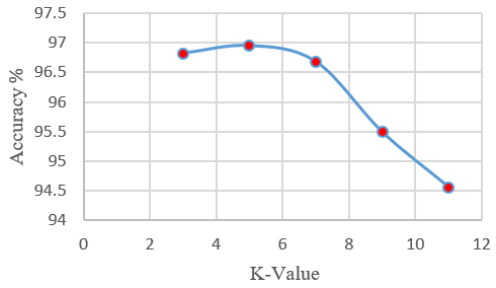

Fig 1.a: Change in accuracy at different k values

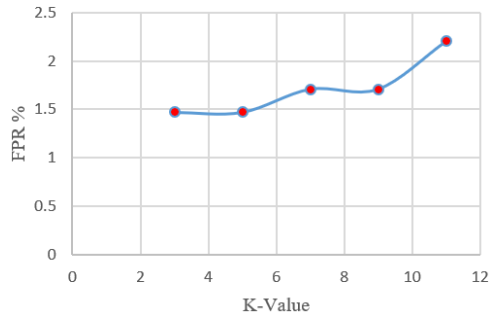

Fig 1.b: Change in FPR at different k values

Fig 1 (a,b): Essessing the classifier performance in Inductive TL approach at different k values.

C. Experimentation of Transductive TL approach

DoS_module is considered as source and R2L_module is considered as target. For the purpose of experimentation all the labels of R2L_module was removed. Fully labeled DoS_module and no labeled R2L_module are given as input to the proposed Transductive TL method for detecting zero-day attacks. KNN classifier is used to classify the zero-day attacks. The best K value is selected as 3 by experimenting on the validation data and the corresponding experimentation process is given in the Fig. 2.

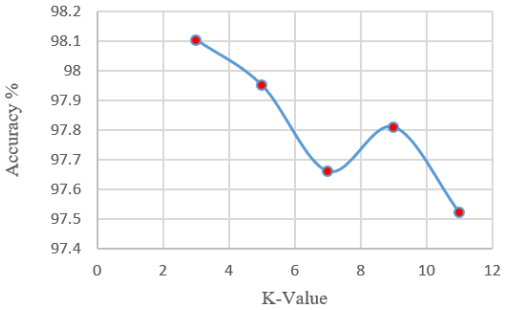

Fig 2.a: Change in accuracy at different k values.

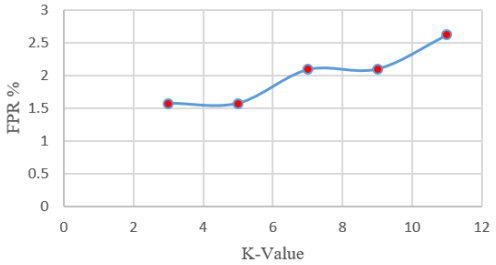

Fig 2.b: Change in FPR at different k values.

Fig 2 (a,b): Essessing the classifier performance in Transductive TL approach at different k values.

The unlabeled target instances are classified by applying KNN with the selected best k value and the proposed approach achieves an accuracy of 75.76% and FPR of 2.62%. The comparative results of both inductive and transductive approaches are given in Table 3.

Table 3: Comparative results of Inductive TL approach and Transductive TL approach

Proposed TL Approach	Source	Target	Accuracy (%)	FPR (%)
Inductive TL	Fully labeled DoS module	10% labeled R2L module	88.24	11.31
Transductive TL	Fully labeled DoS module	Completely unlabeled R2L module	75.76	2.62

From the results it is evident that the Inductive TL approach achieve good accuracy as compared with the Transductive TL approach which indicates that the availability of labeled instances in the target domain leads to the better detection of zero-day attacks.

V. CONCLUSION

Application of transfer learning for IDS problems to detect zero-day attacks is a new research trend followed by modern researchers. This paper explores both inductive and transductive approaches of transfer learning for detecting zero-day attacks. The results suggest that the inductive TL approach achieves better accuracy by utilizing the minimal labeled instances as compared to the transductive TL approach. However, transductive TL due to its attack detection ability even without any labeled instances, is more suitable for zero-day attack detection. Our future research is on enhancing the performance of the transductive TL approach by assigning soft labels to some of the target instances.

REFERENCES

1. Nerella Sameera and M. Shashi, “Intrusion detection analytics: A comprehensive survey”, International Journal of Advanced Scientific Research and Management (IJASRM), Volume 4, Issue 6, ISSN:2455-6378, June, 2019.

2. Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang, "A survey of transfer learning", Journal of Big Data 3.1 (2016): 9.
3. Pan, SinnoJialin, and Qiang Yang, "A survey on transfer learning", IEEE Transactions on knowledge and data engineering 22, no. 10 2010, 1345-1359.
4. Wang, Chang, Sridhar Mahadevan, "Heterogeneous domain adaptation using manifold alignment", Twenty-Second International Joint Conference on Artificial Intelligence, 2011.
5. Nerella Sameera and M Shashi, "Encoding approach for intrusion detection using PCA and KNN classifier", Springer AISC series. to be published.

AUTHORS PROFILE

Mrs. Nerella Sameera has received Bachelor of Technology from Chirala Engineering College, affiliated to JNTU Kakinada in the year 2011 and Master of Technology from Andhra University in year 2013. She is currently pursuing Ph. D in the Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam since 2017. Her main research work focuses on Data Analytics for Cyber Security, Intrusion Detection Systems and Transfer Learning. She has 3 years of teaching experience.

Mrs. Andhavarapu Bhanusri has received Bachelor of Technology from Sri Sivani College of Engineering, affiliated to JNTU Kakinada in the year 2017 and currently pursuing Master of Technology in the department of Computer Science and System Engineering, Andhra University, Visakhapatnam since 2017.

Mrs. M. Shashi received her BE in Electrical and Electronics and ME in Computer Engineering with distinction from Andhra University. She received her PhD in 1994 from Andhra University and obtained her best PhD thesis award. She is working as a Professor of Computer Science and Systems Engineering since 1999 at Andhra University, Visakhapatnam, Andhra Pradesh, India. She received an AICTE career award as young teacher in 1996. She is a co-author of the Indian edition of text book on Data Structures and Program Design in C from Pearson Education Ltd. She published technical papers in national and international journals. Her research interests include data mining, artificial intelligence, pattern recognition and machine learning. She is a member of IEEE, ISTE, CSI and Fellow of Institute of Engineers (India).