ISSN 2395-1621

Distributed Intrusion detection system based on Probabilistic Generative Classifier

[#1]Manisha A.Rakate, [#2]Prof.R.H.Kulkarni

[1]manisharakate@gmail.com
[2]rkpv2002@gmail.com

[#1]BSCOER, Narhe,Pune

ABSTRACT

The security of computer networks plays a strategic role in modern computer systems. In order to enforce high protection levels against threats, a number of software tools are currently developed. Intrusion Detection Systems aim at detecting intruder who eluded the first line protection.

The proposed novel technique of fusing classifiers in Distributed Intrusion detection application with the help of k-nearest neighbor classifier. In this classifier object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.

Keywords— Distributed system, computer network, k-nearest neighbor

ARTICLE INFO

Article History

Received : 16th June 2105

Received in revised form :

17th june 2015

Accepted : 22nd June 2015

Published online :

27th June 2015

I. INTRODUCTION

Data mining applications consist of large amount of data. It is necessary to access this data which is stored in the form of classifiers. Knowledge from two classifiers can't be directly exchange as it may contains large amount of raw data. Therefore, it is impossible to exchange the raw data because of a limited communication bandwidth. Knowledge extraction is split into subtasks due to memory or runtime limitations.[1]

To fuse data of the classifiers we need to derive some model which can probably collect data and combines data from both the classifiers. Probabilistic classifiers are usually follows Bayes theorem which can be formulated as,

$$p(C|X) = \frac{p(X|C)p(c)}{p(X)}$$

x is a random variable which models the input space of the classifier

c is a random variable representing a class (e.g., cϵ{1…,C}).

P(x) is often modelled by means of mixture distributions.

A. Probabilistic Generative Classifier:

Naive Bayes classifier is probabilistic classifier which assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class

Variable [2]. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter.

A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features. An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. Prior distributions must be assumed for various parameters of the classifier. As the components of the classifier already consist of distributions and we now use another set of distributions to describe values of their parameters, the latter set of distributions is called hyper distributions or second-order distributions. The priors are then combined with a likelihood function based on training data.

The hyper distributions of the CMM classifier are trained using a variational inference (VI) algorithm.

II. LITERATURE SURVEY

Fusion can take place at various levels or categories:

1) The classifiers can be used in the form of Ensembles
2) Second, the probabilistic classifiers offer the possibility to combine classifiers at the level of components of the mixture models
3) Components or rules may be fused at the level of parameters

A. First order distribution techniques

1. Sequential Bayesian estimation techniques

Bayesian theory delivers a powerful theoretical platform for the mathematical description and execution of fusion tasks, especially if the information delivering sources are of heterogeneous nature. Fusion tasks increases exponentially with the number of sources [6].

Data fusion design involves making many architectural choices, including: data fusion method, distribution in processing and storage, communication topology, type of exchanged information, degree of pre-processing, and many others. Listing our choices in the same order, the ASN approach to data fusion is: Bayesian decentralized in processing and storage, over a tree or general network, utilizing both scan-to-track and track-to-track fusion, feature-based.

Data or information extracted from the data can be fused to come to more certain conclusion. Bayesian knowledge fusion is associated with this category. Several variants can be found, while the most interesting ones are sequential Bayesian estimation techniques or the fusion of several likelihood functions as in the case of the independent likelihood pool approach [4].

Models trained from sample data can be fused if the models were constructed in a distributed fashion. The output of "low-level" classifiers by averaging their labels or using their labels as input of a decision unit that could also be trained from data. More complex approaches are bagging or boosting which are often motivated by the idea that an ensemble of "weak classifiers" may outperform a single classifier.

2. Variational inference

The basic purpose of Variational methods is finding derivatives of functions. We can think of a function as a mapping that takes the value of a variable as the input and returns the value of the function as the output. The derivative of the function then describes how the output value varies as we make infinitesimal changes to the input value. Similarly, we can define a functional as a mapping that takes a function as the input and that returns the value of the functional as the output [8].

III. EXISTING SYSTEM

Existing system of knowledge fusion makes classification with the help of probabilistic generative classifier. Components or rules may be fused at the level of parameters. In this case, it is necessary to "average" the parameters of two or several components in an appropriate way if these components are regarded as being "sufficiently" similar [3].

This technique is based on the use of probabilistic generative classifiers using multinomial distributions for categorical input dimensions and multivariate normal distributions for the continuous ones.

Disadvantages:

1) Theoretically, naive Bayes classifier have minimum error rate comparing to other classifier, but practically it is not always true, because of assumption of class conditional independence and the lack of available probability data
2) Less accurate compare to other classifier.

IV. PROPOSED SYSTEM

The Problem with Naive Bayes classifier can be solved with the help of K- nearest neighbour classification. K-nearest neighbour is instance based leaning method. Instance based classifiers are also called lazy learners as they store all of the training samples and do not build a classifier until a new, unlabelled sample needs to be classified. Lazy-learning algorithms require less computation time during the training phase than eager-learning algorithms.

The k-nearest neighbours' algorithm is amongst the simplest of all machine learning algorithms. It is based on the principal that the samples that are similar are lies in close proximity [3]. Given an unlabelled sample, K-nearest neighbour classifier searches the pattern space for the k-objects that are closest to it and assigned the class by identifying the most frequent class label. If the value of k=1 then assign the class of the training sample that is the closest to the unknown sample in the pattern space.

Algorithm:

BEGIN

INPUT: D= {(X_1, C_1) , ... , (X_N, C_N)}

X= (X_1 , ... , X_N) new instance to be classified

i. FOR each labelled instance (X_i, C_i) calculate d(X_i, X)
ii. Order d(X_i, X) from lowest to highest , (i=1, ... , N)
iii. Select the K nearest instances to X : D_x^K
iv. Assign to x the most frequent class in D_x^K

Advantages:

1) Easy to understand and easy to implement classification technique.
2) An expected lazy learning methods are faster at training than eager methods.
3) Perform well on application in which a sample can have many class labels

A. System Architecture

Following figure 1 shows the system architecture

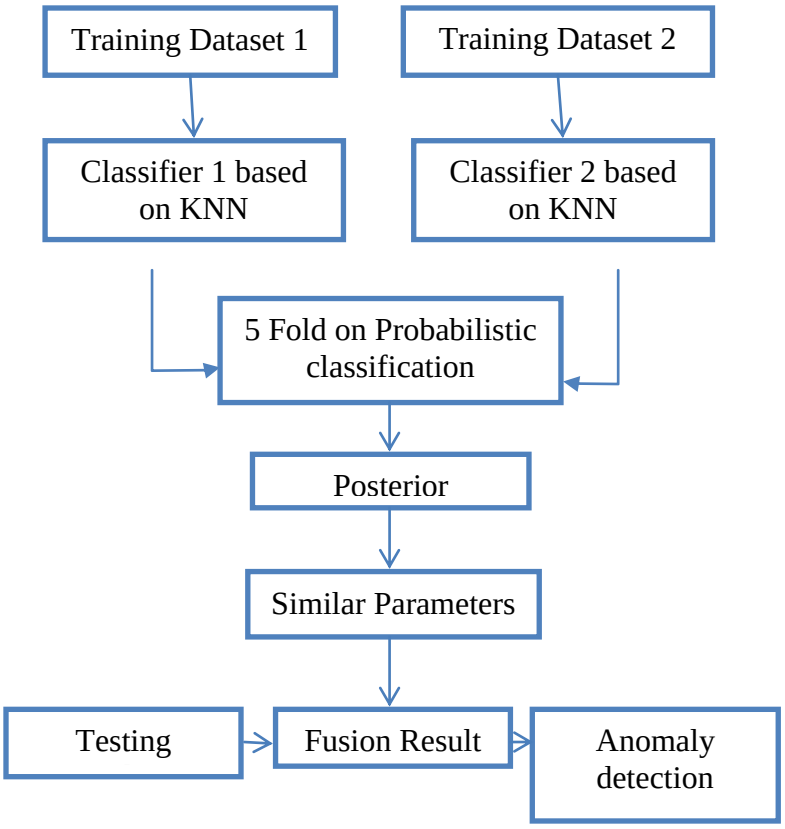

Figure1. System Architecture

B. Architecture Explanation:

1) Components of Training datasets Training dataset 1 and Training dataset 2 are classified with the help of KNN-Classification.
2) For the 2 classifiers apply 5 folds of probabilistic classification.
3) After doing 5 folds of probabilistic classification find posterior classification
4) Final we get fusion result of 2 classifiers.
5) We can further use this classifier for doing anomaly detection.

V. CONCLUSION

Here, a new technology is presented for fusing two probabilistic classifiers into one. CMM subsist of number of components each of which may in turn consist of one multivariate normal distribution modelling continuous dimensions of the input space and multiple multinomial distributions, one for each categorical dimension of the input space. For recognizing the components of two classifiers that shall be fused, a similarity measures recommended which operates on the distribution of the classifier. The genuine fusion of two components works one level higher on the hyper distributions which are obtained from Bayesian training of a CMM using the VI algorithm. Equations to fuse both Dirichlet and normal-Wishart distribution are derived which are the conjugate prior distribution of the multinomial and normal distribution of CMM.

REFRENCES

[1] Dominik Fisch , Edgar Kalkowski " Knowledge fusion for probablistic generative classifier with data mining applications" , IEEE transaction on knowledge and data enginnering, Vol 26, no 3, March 2014

[2] Srinivasa D. Fisch, M. Janicke, E. Kalkowski, and B. Sick, "Learning from Others: Exchange of Classification Rules in Intelligent Distributed Systems," Artificial Intelligence, vol. 187-188, pp. 90-114, 2012

[3] N. Bouguila, "Hybrid Generative/Discriminative Approaches for Proportional Data Modeling and Classification," IEEE Trans. Knowledge and Data Eng., vol. 24, no. 12, pp. 2184-2202, Dec. 2012.

[4] T.M. Hospedales, S. Gong, and T. Xiang, "Finding Rare Classes: Active Learning with Generative and Discriminative Models," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 2, pp. 374-386, Feb. 2013

[5] D. Fisch, T. Gruber, and B. Sick, "SwiftRule: Mining Comprehensible Classification Rules for Time Series Analysis," IEEE Trans. Knowledge and Data Eng., vol. 23, no. 5, pp. 774-787, May 2011.

[6] H. Lu and B. Feng, "An Intelligent Topic Map-Based Approach to Detecting and Resolving Conflicts for Multi-Resource Knowledge Fusion," Information Technology J., vol. 8, no. 8, pp. 1242-1248, 2009.

[7] E. Santos Jr., J. Wilkinson, and E. Santos, "Bayesian Knowledge Fusion," Proc. 22nd Int'l FLAIRS Conf., pp. 559-564, 2009.

[8] K. ying Hui and P. Gray, "Constraint and Data Fusion in a Distributed Information System," Proc. 16th British Nat'l Conf. Databases: Advances in Databases, pp. 181-182, 1998.

[9] H. Durrant-Whyte and T. Henderson, "Multisensor Data Fusion," Springer Handbook of Robotics, B. Siciliano and O. Khatib, eds. chapter 25, pp. 585-610, Springer, 2008.

[10] J. Sander and J. Beyerer, "Fusion Agents—Realizing Bayesian Fusion via a Local Approach," Proc. IEEE Int'l Conf. Multisensor Fusion and Integration for Intelligent Systems, pp. 249-254, 2006.

[11] A.G. Foina, J. Planas, R.M. Badia, and F.J. Ramirez-Fernandez, "Pmeans, A Parallel Clustering Algorithm for a Heterogeneous Multi-Processor Environment," Proc. Int'l Conf. High Performance Computing and Simulation, pp. 239-248, 2011.

[12] C.S.R. Fraser, L.F. Bertuccelli, H.-L. Choi, and J.P. How, "A Hyperparameter- Based Approach for Consensus under Uncertainties," Proc. Am. Control Conf., pp. 3192-3197, 2010.