2009 International Conference on Information and Multimedia Technology

Fusion of Decision Tree and Gaussian Mixture Models for Heterogeneous Data Sets

Khoi-Nguyen Tran
School of Computer Science
The Australian National University
Canberra, Australia
u4315673@anu.edu.au

Huidong Jin
CSIRO Mathematical and Information Sciences
CSIRO, ANU Campus
Canberra, Australia
Warren.Jin@csiro.au

***Abstract*—Current data mining techniques have been developed with great success on homogeneous data. However, few techniques exist for heterogeneous data without further manipulation or consideration of dependencies among the different types of attributes. This paper presents a fusion of C4.5 Decision Tree and Gaussian Mixture Model (GMM) techniques for mixed-attribute data sets. The proposed fusion technique is used to detect anomalies in computer network data. Evaluation experiments were performed on the popular KDDCup 1999 data set using C4.5 Decision Tree, GMM and fusions of C4.5 and GMM. Experimental results showed a better performance for the proposed fusion technique compared to the individual techniques.**

***Keywords*—Fusion technique, C4.5 Decision Tree, Gaussian Mixture Model, Heterogeneous Data, Mixed-Attribute Data, Anomaly Detection, KDDCup 1999.**

I. Introduction

Real-world data sets are normally heterogeneous. For example, network data collected from connections to a computer have a range of symbolic and numeric attributes. In general, data sets have a range of attribute types that must be accounted for and detectors should be flexible enough to adapt to them. Determining a higher meaning to sets of attributes can reduce processing or provide more relevant information to a classifier [1], [2], [3]. However, automated data analysis is usually preferred using methods such as feature selection [4] or principal component analysis [5]. The relationship between the attributes of the data is often overlooked, usually because of the complexity or high dimensionality of the data. In this study, some basic combinations of different data attributes are used to determine if the fusion of the classifiers can make better predictions than using them individually.

Anomaly detection is an important task that has many applications, but current developed methods are usually focused on homogeneous data sets or on converting heterogeneous data to homogeneous data[6], [7], [8]. The mixture of attribute types in network databases causes difficulties when detecting anomalous entries (or "attacks") in these databases. Unsupported attribute types could simply be ignored but this usually results in information being lost from the discarded data. In most of the developed and popular methods, numeric data is converted to symbolic data by means of categorization, or symbolic data is simply enumerated [9]. This data transformation can cause loss of information and may change the meaning of the data. Enumerating data incorrectly gives meaning to the distance between two symbolic data points, and categorisation of data causes information loss of possible data distributions and changes the meaning of the numeric data by assuming groupings that may not exist. Nevertheless, data manipulation often occurs as it simplifies the problem, but it also places assumptions on the data that must be carefully checked [9].

In this paper, we present a fusion technique of C4.5 Decision Tree for symbolic attributes and Gaussian Mixture Model (GMM) for numeric attributes in network data. Evaluation experiments were performed on the popular KDDCup 1999 data set using C4.5 Decision Tree, GMM and the fusion of these two techniques. Experimental results for the individual models and the fusion model showed a better performance for the proposed fusion technique.

The remainder of the paper is organised as follows. Related work is discussed in Section II. A brief discussion on these algorithms is presented in Section III. A description of the proposed fusion model is then given in Section IV, evaluation experiments in Section V, and results are presented in Section VI. Concluding remarks end the paper in Section VII.

II. Related Work

Anomaly detection from mixed-attribute data sets is one of the many challenges of the field, where the focus has been largely on homogenizing the data. However, it is now becoming a primary focus of research for anomaly detection [9]. A close look at anomaly detection using mixed-attribute data is given in [9], where the authors used a distance measure for each of the symbolic and numeric attribute. They also used the KDDCup 1999 data sets in their experiments. For symbolic data, the values define the distance between two records and for numeric data, a covariance matrix is used to analyse how two records are related. An anomaly score is then assigned to each record based on the distance analysis and then classified. Their technique was aimed at a distributed setting, where highly favorable execution times and detection rate were observed.

The work of [9] extends the LOADED and subsequently RELOADED algorithms by [10]. The algorithms focused on finding links between the mixed-attribute data records by assigning anomaly scores to symbolic and numeric attributes

978-0-7695-3922-5/09 $26.00 © 2009 IEEE
DOI 10.1109/ICIMT.2009.59

based on the frequency of symbolic values and a correlation matrix for discretized numeric attributes. The LOADED and RELOADED algorithms also looked at the KDDCup 1999 data set. The algorithms looked at giving scores to each record based on its links with the other records. Improved performance was observed with more lattice levels built from the links. The emphasis of the paper is mainly on the dependencies between the data types and how to find links between two data records based on a similarity measure.

This approach of determining the anomaly score from an analysis of the symbolic and numeric attributes is also seen in the research by [11], where the authors compared their technique to [9]. The symbolic score is based on the frequency of the value in the data set, and the numeric score is based on a cosine function that defines the cosine similarity of numeric attributes. Their technique showed an improvement over [9] in the same KDDCup 1999 data sets for most of the different attack types.

The anomaly detection research above is some of the work that directly addresses mixed-attribute data sets and aims to handle them differently. All the research above have a similar theme of calculating anomalous scores for symbolic and numeric data separately and then classifying those scores based on a defined threshold. The research above also dealt with the KDDCup 1999 data set, but in this study, we aim at improving detection rates of all attacks instead of looking at the performance of each attack type as seen above. The use of anomalous scores in this study came in part from the work mentioned and others, but we determine anomaly scores using classifiers rather than from the data attribute types.

III. Decision Tree and Gaussian Mixture Model

A. Decision Tree Model

Decision trees in the context of machine learning and data mining are predictive models that classify an item based on its characteristics. Specifically, a decision tree is built from a training data set that results in a mapping from the independent variables to a dependent variable. In the context of network intrusion detection, a connection is classified as an attack or normal connection based on the properties (or attributes) of the connection. Decision trees are often used because they are simple to understand and they reflect natural human decision making processes. Decision trees are rules summarized into a tree structure, where based on the conditions defined by the rules, following from the root of the tree to the leaves of the tree determines which label the data record should receive.

Decision trees apply naturally to symbolic attributes, but for numeric attributes, some form of discretizing the values is needed. To find the best or most interesting patterns in the data, the topic of Information Gain from Information Theory has become popular for use with decision trees. Information Gain is used in the C4.5 algorithm, where it is simple and highly effective in generating decision trees. In this paper, the C4.5 Decision Tree algorithm, a popular and powerful classifier developed by Ross Quinlan [12] was used to generate decision rules based on different sets of attributes. The accuracy of the rules generated by C4.5 for the training records is used in the anomaly score of the fusion classifier.

B. Gaussian Mixture Models

Gaussian Mixture Models (GMMs) are used in speech and speaker recognition because of their ability to achieve high accuracy in prediction with the models generated [4]. GMMs are limited to numeric data types so enumeration of symbolic attributes is usually performed. GMMs have also been used in anomaly detection, where [13] surveyed a variety of GMM and probabilistic techniques used for anomaly detection.

GMMs assume that the data points cluster in many places with the center of those clusters called components [14]. Each component has an associated mean vector, where they generate data from a Gaussian with the same mean vector and a covariance matrix. The number of Gaussians is specified to match the number of components, but the number of components is usually not known and so the number of Gaussians specified is usually an estimate. The GMM parameters are iteratively estimated such that they maximize the log-likelihood of the training data using the Expectation-Maximalization (EM) technique [14]. A likelihood score is generated (and used in the fusion as detailed below) for each record of the testing data set and a threshold is specified to determine the anomalous and normal records.

The equations used in this implementation come from [14], but the theoretical basis and much broader descriptions of GMMs can be found in many papers, books and on the Internet. The reader is encouraged to pursue other resources for the theoretical basis of C4.5 and GMMs as the theory cannot be summarized effectively here because of space restrictions.

IV. Fusion Model

The idea of fusion models or more generally known as ensemble models, is to use more than one model for classifying new data on the basis that more models should improve the accuracy as more information is available. In this paper, a weighted linear sum of the scores from the decision tree C4.5 and Gaussian mixture models is used to determine the anomaly score of a data record:

$$S = w_1 * C4.5\_Score + w_2 * GMM\_Score \qquad (1)$$

where $w_1 + w_2 = 1$, $C4.5\_Score \leqslant 1$, $GMM\_Score \leqslant 1$ and thus the combined score $S \leqslant 1$. The $C4.5\_score$ is the accuracy of the rule in the training data set as given by the C4.5 rule generator. The GMM_score is the posterior probability of the data record belonging to the training data set with "normal" labels. The best GMM threshold was determined beforehand (by interval testing) and fixed for the experiments. The initialization of the GMM's posterior probability matrix was with random numbers on every set of results. Many repeated execution of the implementation showed identical sets of results, despite the different resulting GMM weights, means and covariance matrix from training the GMM. That is, all results presented here were identical on multiple runs of the GMM algorithm. This odd observation is likely because of

the narrow range of values in each attribute of the data set. Since the GMM was trained only on the normal connections, we expect that values do not deviate much in the data records of normal connections.

A threshold θ is used to determine how high the combined score S of a record x must be to be classified as a normal connection. If $S < \theta$ then x is intrusive else x is normal. So, the greater the scores are, the greater the chance that the connection is normal. The best threshold observed for the GMM model and fusion model was 0.023, which gives the results presented here.

Some initial experiments with higher and lower weights (that are not severely biased) to each of the classifier scores did not show any interesting changes, where usually only a change in the threshold is needed to show similar results. So, the weights were fixed to be equal (i.e. $w_1 = w_2 = 0.5$ in Equation 1) in this study for simplicity. The weights were introduced with the intention of adding boosting to the fusion model in future experiments.

The fusion model is based on the assumption that the C4.5 and GMM detectors will not always have the same opinions. Using the different information from each classifier, anomalous data could potentially be better distinguished.

V. Data Set and Methodology

A. Data Set Description and Result Presentation Description

The KDDCup 1999 data set was used to evaluate the proposed fusion model [15]. This is a popular data set for developing and evaluating anomaly detection techniques for network traffic. The data set consists of 41 attributes: 7 symbolic and 34 numeric with 4,898,431 record entries for the training data set and 311,029 record entries for the testing data set. The labels of the connections were relabeled in to two classes of "normal" or "attack" connections because of the focus on detecting anomalies.

By using only binary classification, a simple evaluation of the predictive accuracy of the classifiers can be used. The outcome for a binary classification task is usually labeled as positive or negative, where positive labels refer to attacks and negative labels to normal connections. The common method used to evaluate the predictive accuracies of anomaly detectors is through the use of a confusion matrix (or an error matrix), where the predicted label of a record are compared to the actual label.

The focus of intrusion detection is on the True Positive Rate also known as the Detection Rate (DR), the False Positive Rate (FPR) and the overall Accuracy (ACC) of the classifiers. However, it is also useful to look at the Positive Predictive Value (PPV) and the Negative Predictive Value (NPV) because they give an indication of how well the classifier is recognizing attack and normal connections, respectively. These are defined by the following equations from [16]: (# means 'number of')

$$DR = \frac{\#TP}{\#TP+\#FN} * 100\% \qquad FPR = \frac{\#FP}{\#FP+\#TN} * 100\%$$

$$ACC = \frac{\#TP+\#TN}{\#P+\#N} * 100\%$$

$$PPV = \frac{\#TP}{\#TP+\#FP} * 100\% \qquad NPV = \frac{\#TN}{\#TN+\#FN} * 100\%$$

where P is the total positive observations and N is the total negative observations. These measures are common in evaluating anomaly detectors, where [17] discusses the usage of these equations in anomaly detection.

These measures are a method of determining the performance of anomaly detector, which is simple and shows the three important features of a good anomaly detector: high detection rate, low false positive rate, and high accuracy. The error matrix and the accompanying performance measures are summarized in Table I.

TABLE I
Error Matrix with Performance Measures

Actual Class	**Predicted Class**		**Performance**
	Attack	Normal	
Attack	#TP	#FN	DR
Normal	#FP	#TN	FPR
Performance	PPV	NPV	ACC

B. Evaluation Methodology

In our experiments, a C4.5 Decision Tree was built using symbolic attributes and numeric attributes that constitutes basic TCP features as defined in [15], which gives 9 symbolic attributes and 5 numeric attributes. We chose this set of features as it shows a mix of attributes that has a defined meaning of TCP features. We did not choose all or more numeric attributes as the C4.5 algorithm takes a considerably long time to build a tree and the associated rules that we use for the fusion. In contrast, the algorithm did not take longer with more symbolic attributes, so we used all symbolic attributes.

The GMMs were built with all the numeric attributes of the data set. We also trained GMMs on subsets of numeric attributes, but no favorable results were observed; these results are not presented, again because of space. A range of Gaussians (for the components) were tested to determine if adding more Gaussians can improve performance as the features of the data can be better distinguished, as suggested by theory. The number of iterations when training the GMMs was fixed at 10 iterations because little to no improvement in performance was observed for a higher number of iterations. This may be due to the high number of dimensions, which causes a faster convergence because of the sparseness of the data. [18] also observed that higher dimensions also reduced the number of iterations, but different data sets were investigated.

The fusion model investigated is named DTtcpGMM, which was a combination of the C4.5 Decision Tree model DTtcp with the GMMs. The scores from each model were combined as detailed in Section IV.

The C4.5 classifier was trained on all (4,898,431) records and GMMs was only trained on all the normal (approximately 1,000,000) labeled records from the training data set. The classifiers were then applied to a separate testing set consisting of approximately 250,000 attack connections and 60,000 normal connections. These two classifiers were trained differently because C4.5 required at least two classes to build an effective decision tree. This gives two very different detectors, where the decision tree model is built using supervised learning and GMM is built using semi-supervised learning. Any improvements would be seen through the different measures presented.

VI. Results and Discussion

A. C4.5 Decision Tree Model

Table II shows the decision tree classifier applied to the symbolic attributes and the numeric attributes that are basic TCP features (DTtcp) of a connection.

TABLE II
Results of the C4.5 Decision Tree Classifier: DTtcp

Actual Class	Predicted Class		Performance
	Attack	Normal	
Attack	235102	15334	93.88%
Normal	5585	55008	9.22%
Performance	97.68%	78.20%	93.27%

DTtcp shows a favorable detection rate of 93.88%, but the false positive rate is high at 9.22%. So, using DTtcp shows that a high number of attacks are identified, but a high portion of normal connections were incorrectly identified. Looking at the positive and negative predictive values, fewer attacks are recognized, but a lot more normal connections are recognized correctly. These results suggest that relaxing the security measure can result in a higher accuracy and more normal connections correctly classified.

B. Gaussian Mixture Model

Table III shows the performance of the GMM detector with different number of Gaussians, trained on all 34 numeric attributes and the entire training data set. The different number of Gaussians has a small effect on the performance of the classifier with changes less than 1% in the same fields. The classifier is consistent with high detection rates (around 92%), low false positive rate (around 4.9%) and high accuracy (around 92.75%). Similarly high positive predictive value (around 98.7%) and fairly high negative predictive value (around 75%) are also observed. This shows the attack connections are being recognized better than normal connections.

The performance of the GMM detector looks to be, at times, better and worse than DTtcp, where less attacks and more normal connections were correctly classified. So, the GMM detector looks to have comparable performance to decision trees depending on the level of security. Increasing the number of Gaussians does not show a significant improvement in the performance measures as suggested by theory.

TABLE III
Results of the GMM Classifier

Gaussians	Actual Class	Predicted Class		Performance
		Attack	Normal	
4	Attack	230259	20177	91.94%
	Normal	3017	57576	4.98%
	Performance	98.71%	74.05%	92.54%
8	Attack	231302	19134	92.36%
	Normal	2976	57617	4.91%
	Performance	98.73%	75.07%	92.89%
12	Attack	231485	18951	92.43%
	Normal	2964	57629	4.89%
	Performance	98.74%	75.25%	92.95%
16	Attack	230947	19489	92.22%
	Normal	2921	57672	4.82%
	Performance	98.75%	74.74%	92.79%

Overall, the performance of the GMM detector is comparable to the decision tree classifier, where they were trained on different sets of attributes. This shows the effects of using different sets of attributes and classifiers, which can dramatically change the performance of detecting attacks or anomalies in the database.

C. Fusion Model

This section will present a fusion of the decision tree and GMM detector: DTtcp and GMM. The fusion model was built based on the description given in Section IV. The thresholds of the fusion models were determined similarly to the GMM detector, where the results were optimal after a certain threshold.

TABLE IV
Results of the Fusion Classifier: DTtcpGMM

Gaussians	Actual Class	Predicted Class		Performance
		Attack	Normal	
4	Attack	235066	15370	93.86%
	Normal	1456	59137	2.40%
	Performance	99.38%	79.37%	94.59%
8	Attack	235071	15365	93.86%
	Normal	1454	59139	2.40%
	Performance	99.39%	79.38%	94.59%
12	Attack	235074	15362	93.86%
	Normal	1454	59139	2.40%
	Performance	99.39%	79.38%	94.59%
16	Attack	235076	15360	93.87%
	Normal	1550	59043	2.56%
	Performance	99.34%	79.36%	94.56%

Table IV shows the results of the fusion of DTtcp and GMM. The results of DTtcpGMM show good performance measures on all Gaussians measured. Significant improvements are observed in the detection rate of over 1.5% compared to GMM in most cases, and a much lower false positive rate of over 2.5% in most cases. Consequently, this resulted in an accuracy that is higher than all the other classifiers by at least 1.5%. Similarly, the positive and negative predictive values for the DTtcpGMM are also very high, suggesting DTtcpGMM is correctly recognizing more normal and attack connections.

To determine that the differences between the GMM classifier and its fusion are significant, a (two-tailed) paired t-test is presented. An extremely statistically significant p-value of 0.0003 is obtained for the set accuracy of DTtcpGMM shown in Table IV and set of accuracy of GMM in Table III. A (two-tailed) paired t-test between the accuracy of DTtcpGMM and DTtcp gives a p-value of 0.0001, which is also extremely statistically significant. (Recall that the algorithms were executed multiple times with the same results.)

So, the performance of DTtcpGMM is statistically better than its individual parts. The small overfitting observed from DTtcp may have helped in distinguishing normal and attack connections when combined with the GMM detector. The dependencies of the attributes to each other look to be very complex as the introduction of some numeric attributes to the decision tree showed signs of overfitting, but when combined with another classifier like the GMM, improvements greater than both those classifiers are seen.

Overall, the fusion of DTtcp and GMM shows a greater improvement than their individual classifiers. The fusion also shows that there is little to no improvements in performance with higher Gaussians, so fewer Gaussians could be used to reduce the training time. The small overfitting of the decision tree classifier (i.e. high FPR) does not poorly influence the classifier as expected, but better performance is observed when combined with GMM.

VII. Conclusion and Future Research

This paper investigated a fusion of the C4.5 Decision Tree classifier and Gaussian Mixture Models for anomaly detection on a mixed-attribute data set, the KDDCup 1999 data set. The C4.5 Decision Tree algorithm was evaluated on symbolic and numeric attributes that constitutes a TCP connection, and the Gaussian Mixture Model was trained on the numeric attributes only. We focused on applying these two techniques to heterogeneous data without manipulating the data to suit one particular technique. A new fusion technique was proposed and evaluated with a series of experiments that showed statistically significant improvements in detection performance compared to the two individual techniques.

From this study, there are many other possible investigation paths, but automatic feature selection [4] may likely prove to be beneficial to choosing the best sets of attributes for the different classifiers. Other data sets also need to be considered because of the age and problems with the KDDCup 1999 data set as strongly criticized by [19], but we are not aware of alternatives at this time. Using more classifiers in the fusion may also show additional benefits because "in reality there are many different types of intrusions, and different detectors are needed to detect them." [20]

References

[1] D. Marchette, "A statistical method for profiling network traffic", Proceedings of the 1st conference on Workshop on Intrusion Detection and Network Monitoring, pp. 13-17, 1999.

[2] C. C. Aggarwal, "Re-designing distance functions and distance-based applications for high dimensional data", SIGMOD Rec., vol. 30, no. 1, pp. 13-18, 2001.

[3] P. Guo, J-Y. Dai and Y-X. Wang, "Outlier Detection in High Dimension Based on Projection", International Conference on Machine Learning and Cybernetics, pp. 1165-1169, 2006.

[4] D. Tran, W. Ma, and D. Sharma, "Automated network feature weighting-based anomaly detection", IEEE International Conference on Intelligence and Security Informatics, pp. 162-166, 2008.

[5] M. Tavallaee, W. Lu, S. A. Iqbal and A. A. Ghorbani, "A Novel Covariance Matrix Based Approach for Detecting Network Anomalies", Proceedings of the Communication Networks and Services Research Conference, pp. 75-81, 2008.

[6] P.K. Chan, M.V. Mahoney and M.H. Arshad, "A machine learning approach to anomaly detection", Technical Report, Department of Computer Science, Florida, Institute of Technology, 2003.

[7] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn and L.-W.Chang "A novel anomaly detection scheme based on principal component classifier", Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop, 2003.

[8] H. Yang, F. Xie, and Y. Lu, "Clustering and Classification Based Anomaly Detection", Lectures Notes in Computer Science, vol. 4223, pp. 1082-1091, 2006.

[9] M. E. Otey, A. Ghoting and S. Parthasarathy, "Fast Distributed Outlier Detection in Mixed-Attribute Data Sets", Data Min. Knowl. Discov., vol. 12, no. 2-3, pp. 203-228, 2006.

[10] A. Ghoting, M.E. Otey and S. Parthasarathy, "LOADED: Link-Based Outlier and Anomaly Detection in Evolving Data Sets", in Proceedings of the Fourth IEEE International Conference on Data Mining, pp. 387-390, 2004.

[11] A. Koufakou, M. Georgiopoulos and G.C. Anagnostopoulos, "Detecting Outliers in High-Dimensional Datasets with Mixed Attributes", Proceedings of DMIN, pp. 427-433, 2008.

[12] R. Quinlan, "C4.5 Release 8", http://www.rulequest.com/Personal/.

[13] M. Markou and S. Singh, "Novelty detection: a review—Part 1: statistical approaches", Signal Processing, vol. 83, no. 12, pp. 2481-2497, Elsevier, 2003.

[14] D. Tran, M. Wagner, and T.V. Le, "Fuzzy Gaussian mixture models for speaker recognition", Proceedings of the International Conference on Spoken Language Processing, pp. 759-762, 1998.

[15] KDD Cup 1999 Data, UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/

[16] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers", Technical Report, HP Laboratories, California, 2004.

[17] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends", Elsevier Computer Networks, vol. 51, issue 12, pp. 34483470, 2007.

[18] Z. Zhang, B.T.Dai, and A. K. H. Tung, "Estimating local optimums in em algorithm over gaussian mixture model", Proceedings of the 25th international conference on Machine learning, pp. 12401247, 2008.

[19] Mc Hugh, "Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by lincoln laboratory", ACM Transactions on Information and System Security 3, 4 (November), pp. 262294, 2000.

[20] S. Axelsson and D. Sands, "An Introduction to Intrusion Detection", in Understanding Intrusion Detection Through Visualization, Springer Verlag, pp. 15-29, 2006.

[21] Z.-S. Pan, S.-C. Chen, G.-B. Hu, and D.-Q. Zhang, "Hybrid neural network and c4.5 for misuse detection", International Conference on Machine Learning and Cybernetics, vol. 4, pp. 24632467, 2003

[22] Y. Freund and R.E. Schapire, "A brief introduction to boosting", Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, pp. 14011406, 1999.

[23] W. Hu and W. Hu, "Network-based intrusion detection using adaboost algorithm", Proceedings of The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 712717, 2005.

[24] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection : A Survey", to Appear in ACM Computing Surveys, 2009.

[25] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach", Prentice Hall, 2003.

[26] H.-D. Jin, M.-L. Wong, and K.-S. Leung, "Scalable Model-based Clustering for Large Databases Based on Data Summarization", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no 11, pp. 1710-1719, 2005.

[27] M. Bahrololum and M. Khaleghi, "Anomaly Intrusion Detection System Using Hierarchical Gaussian Mixture Model", International Journal of Computer Science and Network Security, vol. 8, no. 8, pp. 264-271, 2008.