Classification Based Outlier Detection Techniques

Dr. Shuchita Upadhyaya, Karanjit Singh
Dept. of Computer Science and Applications, Kurukshetra University
Haryana, India
HQ Base Workshop Group EME, Meerut Cantt
UP, India

***Abstract* — Outlier detection is an important research area forming part of many application domains. Specific application domains call for specific detection techniques, while the more generic ones can be applied in a large number of scenarios with good results. This survey tries to provide a structured and comprehensive overview of the research on Classification Based Outlier Detection listing out various techniques as applicable to our area of research. We have focused on the underlying approach adopted by each technique. We have identified key assumptions, which are used by the techniques to differentiate between normal and Outlier behavior. When applying a given technique to a particular domain, these assumptions can be used as guidelines to assess the effectiveness of the technique in that domain. We provide a basic outlier detection technique, and then show how the different existing techniques in that category are variants of this basic technique. This template provides an easier and succinct understanding of the Classification based techniques. Further we identify the advantages and disadvantages of various classification based techniques. We also provide a discussion on the computational complexity of the techniques since it is an important issue in our application domain. We hope that this survey will provide a better understanding of the different directions in which research has been done on this topic, and how techniques developed in this area can be applied in other domains for which they were not intended to begin with.**

***Keywords* — Outliers, Classification, Outlier Detection, Classification based Outlier Detection, One-Class, Multi-Class, Algorithms, Data Mining.**

I. INTRODUCTION

A. *General Description and Underlying Assumptions*

Classification [1,2] is used to learn a model (classifier) from a set of labeled data instances (*training*) and then, classify a test instance into one of the classes using the learnt model (*testing*). Classification based outlier detection techniques operate under the general assumption that a classifier can be learnt from a given feature space that can distinguish between normal and outlier classes.

B. *General Methodology of Operation*

Classification based outlier detection techniques operate in two phases.

1) Training - The training phase learns a classifier using the available labeled training data.

2) Testing – This phase classifies a test instance as normal or an outlier using the classifier.

II. CATEGORISATION OF CLASSIFICATION BASED OUTLIER DETECTION TECHNIQUES

These techniques can be grouped into two categories:-

A. *Multi-Class*

These techniques assume that the training data contains labeled instances which belong to multiple normal classes [3,4]. One has to learn a classifier to distinguish between each normal class against the rest of the classes. Refer Figure 1 for illustration. If a test instance is not classified as normal by any of the classifiers, then it is considered as an outlier. Some multi-category techniques give the prediction made by the classifier a confidence score. If none of the classifiers are confident in classifying the test instance as normal or in other words do not score well, the instance is declared to be an outlier.

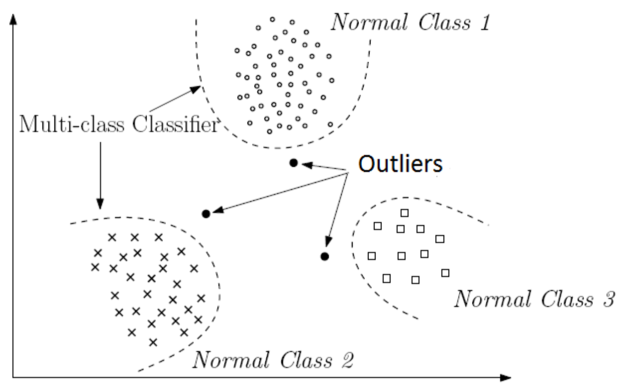

Figure 1: Multi-class Classification Based Outlier Detection

B. *One-Class*

One-class classification based outlier detection techniques assume that all training instances bear a single class label. After learning a discriminative boundary around the normal instances using suitable one-class classification algorithm if any test instance does not fall within the learnt boundary it is considered an outlier. Some well known algorithms are as listed below:-

1) *One-class SVMs* [5].

2) *One-class Kernel Fisher Discriminants* [6,7], as shown in Figure 2.

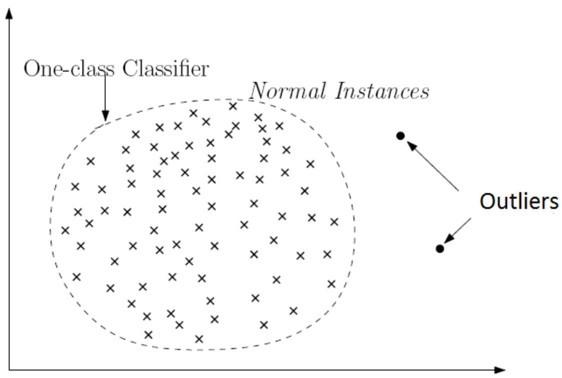

Figure 2 : One-class Classification Based Outlier Detection

III. NEURAL NETWORKS BASED TECHNIQUES

A. *General Description*

Neural networks can be applied to outlier detection in multi-class as well as one-class scenarios. The basic technique of operation of multi-class outlier detection using neural networks operates is again two step.

1) *Training :* First, in the training phase the neural network is trained on normal training data in order to learn the different normal classes.

2) *Testing :* Second, wherein each test instance is provided as an input to the neural network. If the test input is accepted by the network, it is considered normal and if the network rejects a test input, it is considered as an outlier [3,8].

B. *Variants of the Technique*

Several variants of the basic neural network technique have been proposed that use different types of neural networks, as summarized in Table I. These are elaborated subsequently.

TABLE I
SOME REFERENCES CLASSIFICATION BASED OUTLIER DETECTION TECHNIQUES USING NEURAL NETWORKS

Neural Network Used	References
Multi Layered Perceptrons	[9,10,11,12]
Radial Basis Function Based	[13,14,15]
Neural Trees	[16]
Oscillatory Networks	[17,18]
Auto-associative Networks	[19,20,21,22]
Adaptive Resonance Theory Based	[23,24,25]
Hopfield Networks	[26,27,28]

C. *Replicator Neural Networks*

These can be used for one-class outlier detection [Hawkins et al. 2002; Williams et al. 2002]. Using their inputs as outputs internally, they are self organizing, forming a compressed representation for the input data. Usually a multi-layer feed forward neural network is constructed that has the same number of input and output neurons (corresponding to the features in the data). In the training phase data is compressed into three hidden layers. The testing phase involves reconstructing each data instance x_i using the learnt network to obtain the reconstructed output o_i. The reconstruction error for the test instance x_i is then computed as

$$\delta_i = \frac{1}{n}\sum_{j=1}^{n}(x_{ij} - o_{ij})^2$$

where n is the number of features over which the data is defined. The reconstruction error is directly used as an outlier score for the test instance.

IV. BAYESIAN NETWORKS BASED TECHNIQUES

A. *General Description for Univariate Data Set*

Bayesian networks can be used in the multi-class setting for outlier detection. In its basic form for a univariate categorical data set, a naïve Bayesian network estimates the posterior probability of observing a class label (from a set of normal class labels and the outlier class label), given a test data instance. The predicted class for the given test instance is the class label with the largest posterior. The likelihood of observing the test instance, given a class and the prior on the class probabilities, are estimated from the training data set. Laplace Smoothing is used for smoothing the zero probabilities, especially for the outlier class.

B. *Extension of the Technique to Multivariate Data*

By aggregating the per-attribute posterior probabilities for each test instance and then using this aggregated value to assign a class label to the test instance the same basic technique can be generalized for multivariate categorical data set. The basic technique described above assumes independence between the different attributes. Examples are as tabulated below:-

TABLE II

OUTLIER DETECTION TECHNIQUES USING BAYESIAN NETWORKS FROM MULTIVARIATE DATA

Application	References
Network Intrusion Detection	[29,30]
Novelty Detection in Video Surveillance	[31]
Outlier Detection in Text Data	[32]
Disease Outbreak Detection	[33]

More complex Bayesian networks such as those proposed in [34, 35, 36] capture the conditional dependencies between the different attributes.

V. Support Vector Machines Based

A. General Description

Support Vector Machines (SVMs) [37] have been applied to outlier detection in the one-class setting. Such techniques use one class learning [38] to learn a region that contains the training data instances (a boundary). For learning complex regions Kernels, such as *radial basis function (RBF) kernel*, can be used. For each test instance, the basic technique determines if the test instance falls within the learnt region. If a test instance falls within the learnt region, it is declared as normal, else it is declared as an outlier.

B. Variants

TABLE III

OUTLIER DETECTION USING SUPPORT VECTOR MACHINE TECHNIQUES FOR MULTIVARIATE DATA

Application	**References**
Outlier Detection in Audio Signal Data	[39]
Novelty Detection in Power Generation Plants	[40]
System Call Intrusion Detection	[41]
Outliers in Temporal Sequences	[42]

A variant of the basic technique [43] finds the smallest hyper-sphere in the kernel space, which contains all training instances, and then determines which side of that hyper-sphere does a test instance lie. If a test instance lies outside the hyper-sphere, it is declared to be an outlier.

Robust Support Vector Machines (RSVM) which are robust to the presence of outliers in the training data have been proposed by [44]. RSVM have also been applied to system call intrusion detection [45].

VI. Rule Based ASSOCIATION

C. General Description

Rule based outlier detection techniques learn rules that capture the normal behaviour of a system. A test instance not covered by any such rule is considered as an outlier. Such techniques can be applied in one-class as well as multi-class setting.

D. Basic Technique

A basic multi-class rule based technique consists of two steps.

1) Learning Rules from Training Data - Each rule has an associated confidence value. This is proportional to the ratio between the number of training instances correctly classified by the rule and total number of training instances covered by the rule. Algorithms, such as RIPPER, Decision Trees are commonly used for learning.

2) For Each Test Instance Find the Rule that Best Captures the Test Instance - The inverse of the confidence associated with the best rule is the outlier score of the test instance.

TABLE IV

ASSOCIATION RULE BASED OUTLIER DETECTION TECHNIQUES USING MULTIVARIATE DATA

Application	**References**
Network Intrusion Detection	[46,47]
System Call Intrusion Detection	[48, 49]
Credit Card Fraud Detection	[51]
Fraud Detection in Spacecraft House Keeping Data	[52]

Association rules are generated from a categorical data set. To ensure that the rules correspond to strong patterns, a support threshold is used to prune out rules with low support [53]. Association rule mining [54] has been used for one-class outlier detection by generating rules from the data in an unsupervised fashion. In the intermediate step of association rule mining algorithms, frequent item sets are generated. An outlier detection algorithm for categorical data sets has been proposed in which the outlier score of a test instance is equal to the number of frequent item sets it occurs in [54].

VII. Computational Complexity

The computational complexity of classification based techniques depends on the classification algorithm being used.

A. Training Phase

The complexity of training classifiers has been covered in [56]. Generally, training decision trees tends to be faster while techniques that involve quadratic optimization, such as SVMs, are more expensive, though some linear time SVMs [57] have been proposed that have linear training time.

B. Testing Phase

The testing phase of classification techniques is usually very fast since the testing phase uses a learnt model for classification.

VIII. Advantages and Disadvantages of Classification Based Techniques

A. Advantages

The advantages of classification based techniques are as follows:

1) Use of Powerful Algorithms - Classification based techniques, especially multi-class, can make use of powerful algorithms that can distinguish between instances belonging to different classes.

2) Fast testing Phase - The testing phase is fast since each test instance gets compared against the pre-computed model.

B. Disadvantages

The disadvantages of classification based techniques are as follows:

1) Non-Availability of Accurate Labels for Various Normal classes - Multi-class classification based techniques rely on availability of accurate labels for various normal classes, which is often not possible.

2) Assigning Label to Each Test Instance – This can become a disadvantage when a meaningful outlier score is desired for the test instances being subject to classification based techniques. However some classification techniques that obtain a probabilistic prediction score from the output of a classifier, can be used to address this issue [58].

IX. CONCLUDING REMARKS AND FUTURE WORK

In this survey we have discussed different ways in which the problem of classification based outlier detection has been formulated in literature, and we attempted to provide an overview of the huge literature on different techniques. For each subcategory of Classification based technique, we could identify a unique assumption regarding the notion of normal data and outliers. When applying a given technique to a particular domain, these assumptions can be used as guidelines to assess the effectiveness of the technique in that domain. We understand that ideally, a comprehensive survey should not only allow a reader to understand the motivation behind using a particular technique, but also provide a comparative analysis of various techniques. But the current research done in an unstructured fashion, without relying on a unified notion of outliers, makes a theoretical understanding of the outlier detection problem a difficult task. A possible future work would be to unify the assumptions made by different techniques regarding the normal and outlier behavior into a statistical or machine learning framework.

X. ACKNOWLEDGMENTS

Dr Shuchita Upadhyaya and Karanjit Singh and thanks the staff of Computer Science and Applications department of Kurukshetra University for their wholehearted support in referencing the study material. The authors sincerely thank the library staff for the late duty hours at times.

REFERENCES

[1] Tan, P.-N., Steinbach, M., and Kumar, V. 2005. Introduction to Data Mining. Addison-Wesley.

[2] Duda, R. O., Hart, P. E., and Stork, D. G. 2000. Pattern Classification (2nd Edition). Wiley-Interscience.

[3] Stefano, C., Sansone, C., and Vento, M. 2000. To reject or not to reject: that is the question - an answer in case of neural classifiers. IEEE Transactions on Systems, Management and Cybernetics 30, 1, 84 - 94.

[4] Barbara, D., Couto, J., Jajodia, S., and Wu, N. 2001b. Detecting novel network intrusions using bayes estimators. In Proceedings of the First SIAM International Conference on Data Mining.

[5] Scholkopf,Ä B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., and Williamson, R. C. 2001. Estimating the support of a high-dimensional distribution. Neural Comput. 13, 7, 1443 - 1471.

[6] Roth, V. 2004. Outlier detection with one-class kernel fisher discriminants.

[7] Roth, V. 2006. Kernel fisher discriminants for outlier detection. Neural Computation 18, 4, 942 - 960.

[8] Odin, T. and Addison, D. 2000. Novelty detection using neural network technology. In Proceedings of the COMADEN Conference. Houston, TX.

[9] Ghosh, A. K., Schwartzbard, A., and Schatz, M. 1999a. Learning program behavior profiles for intrusion detection. In Proceedings of 1st USENIX Workshop on Intrusion Detection and Network Monitoring. 51 - 62.

[10] Ghosh, A. K., Wanken, J., and Charron, F. 1998. Detecting anomalous and unknown intrusions against programs. In Proceedings of the 14th Annual Computer Security Applications Conference. IEEE Computer Society, 259.

[11] Barson, P., Davey, N., Field, S. D. H., Frank, R. J., and McAskie, G. 1996. The detection of fraud in mobile phone networks. Neural Network World 6, 4.

[12] Hickinbotham, S. J. and Austin, J. 2000b. Novelty detection in airframe strain data. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. Vol. 6. 24 - 27.

[13] Bishop, C. 1994. Novelty detection and neural network validation. In Proceedings of IEEE Vision, Image and Signal Processing. Vol. 141. 217 - 222.

[14] Ghosh, S. and Reilly, D. L. 1994. Credit card fraud detection with a neural-network. In Proceedings of the 27th Annual Hawaii International Conference on System Science. Vol. 3. Los Alamitos, CA.

[15] Jakubek, S. and Strasser, T. 2002. Fault-diagnosis using neural networks with ellipsoidal basis functions. In Proceedings of the American Control Conference. Vol. 5. 3846 - 3851.

[16] Martinez, D. 1998. Neural tree density estimation for novelty detection. IEEE Transactions on Neural Networks 9, 2, 330 - 338

[17] Ho, T. V. and Rouat, J. 1997. A novelty detector using a network of integrate and fire neurons. Lecture Notes in Computer Science 1327, 103 - 108.

[18] Aeyels, D. 1991. On the dynamic behaviour of the novelty detector and the novelty filter. In Analysis of Controlled Dynamical Systems- Progress in Systems and Control Theory, B. Bon-nard, B. Bride, J. Gauthier, and I. Kupka, Eds. Vol. 8. Springer, Berlin, 1 - 10.

[19] Hawkins, S., He, H., Williams, G. J., and Baxter, R. A. 2002. Outlier detection using replicator neural networks. In Proceedings of the 4th International Conference on Data Warehousing and Knowledge Discovery. Springer-Verlag, 170 - 180.

[20] Song, S., Shin, D., and Yoon, E. 2001. Analysis of novelty detection properties of auto-associators. In Proceedings of Condition Monitoring and Diagnostic Engineering Management. 577 - 584.

[21] Streifel, R., Maks, R., and El-Sharkawi, M. 1996. Detection of shorted-turns in the field of turbine-generator rotors using novelty detectors - development and field tests. IEEE Transactions on Energy Conversations 11, 2, 312 - 317.

[22] Worden, K. 1997. Structural fault detection using a novelty measure. Journal of Sound Vibration 201, 1, 85 - 101.

[23] Moya, M., Koch, M., and Hostetler, L. 1993. One-class classifier networks for target recognition applications. In Proceedings on World Congress on Neural Networks, International Neural Network Society. Portland, OR, 797 - 801.

[24] Dasgupta, D. and Nino, F. 2000. A comparison of negative and positive selection algorithms in novel pattern detection. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Vol. 1. Nashville, TN, 125 - 130.

[25] Caudell, T. and Newman, D. 1993. An adaptive resonance architecture to define normality and detect novelties in time series and databases. In IEEE World Congress on Neural Networks. IEEE, Portland, OR, 166 - 176.

[26] Jagota, A. 1991. Novelty detection on a very large number of memories stored in a hopfield-style network. In Proceedings of the International Joint Conference on Neural Networks. Vol. 2. Seattle, WA, 905.

[27] Addison, J., Wermter, S., and MacIntyre, J. 1999. Effectiveness of feature extraction in neural network architectures for novelty detection. In Proceedings of the 9th International Conference on Artificial Neural Networks. Vol. 2. 976 - 981.

[28] Murray, A. F. 2001. Novelty detection using products of simple experts - a potential architecture for embedded systems. Neural Networks 14, 9, 1257 - 1264.

[29] Sebyala, A. A., Olukemi, T., and Sacks, L. 2002. Active platform security through intrusion detection using naive bayesian network for outlier detection. In Proceedings of the 2002 London Communications Symposium.

[30] Bronstein, A., Das, J., Duro, M., Friedrich, R., Kleyner, G., Mueller, M., Singhal, S., and Cohen, I. 2001. Bayesian networks for detecting outliers in internet-based services. In International Symposium on Integrated Network Management.

[31] Diehl, C. and Hampshire, J. 2002. Real-time object classification and novelty detection for collaborative video surveillance. In Proceedings of IEEE International Joint Conference on Neural Networks. IEEE, Honolulu, HI.

[32] Baker, D., Hofmann, T., McCallum, A., and Yang, Y. 1999. A hierarchical probabilistic model for novelty detection in text. In Proceedings of International Conference on Machine Learning.

[33] Wong, W.-K., Moore, A., Cooper, G., and Wagner, M. 2002. Rule-based outlier pattern detection for detecting disease outbreaks. Available online from http://www.cs.cmu.edu/simawm/antiterror.

[34] Siaterlis, C. and Maglaris, B. 2004. Towards multisensor data fusion for dos detection. In Proceedings of the 2004 ACM symposium on Applied computing. ACM Press, 439 - 446.

[35] Janakiram, D., Reddy, V., and Kumar, A. 2006. Outlier detection in wireless sensor networks using bayesian belief networks. In First International Conference on Communication System Software and Middleware. 1 - 6.

[36] Das, K. and Schneider, J. 2007. Detecting anomalous records in categorical datasets. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM Press.

[37] Vapnik, V. N. 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA.

[38] Ratsch, G., Mika, S., Scholkopf, B., and Muller, K.-R. 2002. Constructing boosting algo-rithms from svms: An application to one-class classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 9, 1184 - 1199

[39] Davy, M. and Godsill, S. 2002. Detection of abrupt spectral changes using support vector machines. an application to audio signal segmentation. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, USA.

[40] King, S., King, D., P. Anuzis, K. A., Tarassenko, L., Hayton, P., and Utete, S. 2002. The use of novelty detection techniques for monitoring high-integrity plant. In Proceedings of the 2002 International Conference on Control Applications. Vol. 1. Cancun, Mexico, 221 - 226.

[41] Heller, K. A., Svore, K. M., Keromytis, A. D., and Stolfo, S. J. 2003. One class support vector machines for detecting anomalous windows registry accesses. In Proceedings of the Workshop on Data Mining for Computer Security.

[42] Ma, J. and Perkins, S. 2003a. Online novelty detection on temporal sequences. In Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM Press, New York, NY, USA, 613 - 618.

[43] Tax, D. and Duin, R. 1999a. Data domain description using support vectors. In Proceedings of the European Symposium on Artificial Neural Networks, M. Verleysen, Ed. Brussels, 251 - 256.

[44] Song, Q., Hu, W., and Xie, W. 2002. Robust support vector machine with bullet hole image classification. IEEE Transactions on Systems, Man, and Cybernetics - Part C:Applications and Reviews 32, 4.

[45] Hu, W., Liao, Y., and Vemuri, V. R. 2003. Robust outlier detection using support vec-tor machines. In Proceedings of the International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 282 - 289.

[46] Mahoney, M. V. and Chan, P. K. 2002. Learning nonstationary models of normal network traffic for detecting novel attacks. In Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM Press, 376 - 385

[47] Mahoney, M. V. and Chan, P. K. 2003. Learning rules for outlier detection of hostile network traffic. In Proceedings of the 3rd IEEE International Conference on Data Mining. IEEE Computer Society, 601.

[48] Lee, W., Stolfo, S. J., and Mok, K. W. 2000. Adaptive intrusion detection: A data mining approach. Artificial Intelligence Review 14, 6, 533 - 567.

[49] Lee, W. and Stolfo, S. 1998. Data mining approaches for intrusion detection. In Proceedings of the 7th USENIX Security Symposium. San Antonio, TX.

[50] Qin, M. and Hwang, K. 2004. Frequent episode rules for internet outlier detection. In Proceedings of the 3rd IEEE International Symposium on Network Computing and Applications. IEEE Computer Society.

[51] Brause, R., Langsdorf, T., and Hepp, M. 1999. Neural data mining for credit card fraud detection. In Proceedings of IEEE International Conference on Tools with Artificial Intelligence. 103 - 106

[52] Yairi, T., Kato, Y., and Hori, K. 2001. Fault detection by mining association rules from house-keeping data. In In Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space.

[53] Tan, P.-N., Steinbach, M., and Kumar, V. 2005. Introduction to Data Mining. Addison-Wesley.

[54] Agrawal, R. and Srikant, R. 1995. Mining sequential patterns. In Proceedings of the 11th International Conference on Data Engineering. IEEE Computer Society, Washington, DC, USA, 3 - 14.

[55] He, Z., Xu, X., Huang, J. Z., and Deng, S. 2004a. A frequent pattern discovery method for outlier detection. 726 - 732.

[56] Kearns, M. J. 1990. Computational Complexity of Machine Learning. MIT Press, Cambridge, MA, USA.

[57] Joachims, T. 2006. Training linear svms in linear time. In KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA, 217 - 226.

[58] Platt, J. 2000. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. A. Smola, P. Bartlett, B. Schoelkopf, and D. Schuurmans, Eds. 61 - 74