International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Online Anomaly Detection under Over-sampling PCA

Y Srilakshmi[1], D Ratna Kishore[2]

[1]M.Tech, Department of Computer Science & Engineering,
Dhanekula Institute of Engineering & Technology, India

[2]Assistant Professor, Department of Computer Science & Engineering,
Dhanekula Institute of Engineering & Technology, JNTU Kakinada, India

Abstract: ***Anomaly detection is the process of identifying unusual behavior. Outlier detection is an important issue in data mining and has been studied in different research areas. In this paper we use "Leave One Out" procedure to check each individual point the "with or without" effect on the variation of principal directions. Based on this idea, an over-sampling principal component analysis (osPCA) outlier detection method is proposed for emphasizing the influence of an abnormal instance. Except for identifying the suspicious outliers, we also design an online anomaly detection to detect the new arriving anomaly. In addition, we also study the quick updating of the principal directions for the effective computation and satisfying the online detecting demand. It is widely used in data mining; the proposed framework is favored for online applications which have computation or memory limitations. Compared with the all existing algorithms, our proposed method is in terms of flexibility, accuracy and efficiency.***

Keywords: over sampling, anomaly detection, fault detection, Leave One Out, Principle component analysis.

1. Introduction

Anomaly detection has been an important research topic and it is widely used in data mining. Many real world applications such as intrusion or credit card fraud detection require an effective framework to identify deviated data instances. Due to the reasons that only few labeled data are available in applications and the events that people are interested in are extremely rare or do not happen before, the outlier detection is getting people's attention more and more [1,2,3,4,5,6]. Thus the outlier detection methods are designed for finding the rare instances or the deviated data. In other words, an outlier detection method can be applied to deal with exists in a small proportion of network traffic.

However the rareness of the deviated data, its presence might enormously affect the solution model such as the distribution or principle direction of the data. In the past, many outlier detection methods have been proposed [1, 3, 5]. One of the most popular methods is using the density-based local outlier factor (LOF) to measure the outlierness for each instance [1]. The LOF uses the density of each individual instance's neighbors to define the degree of outlierness and concludes a suspicious ranking for all instances. The most important property of the LOF is considering the local data structure for eliminating the density. In general all data instances have to be compared with all other remaining ones leading to computational complexity of **O (N^2d),** whereas d is the dimension of the data. This property makes the LOF discover the outliers which are sheltered under a global data structure.

Besides, angle-based outlier detection (ABOD) method has also been proposed recently [5]. The main concept of ABOD is using the variation of the angles between the each target instance and the rest instances. An outlier or deviated instance will generate a smaller variance among its associated angles. Based on this observation, the ABOD considers all the variance of angles between the target instance and any pair of instances to detect outliers. In [5], the authors also proposed the fast ABOD which is an approximation of the original ABOD. The difference is that fast ABOD only considers the variance of the angles between the target instance and any pair of instances of target instance's *k* nearest neighbors. Even though, these methods mentioned above cannot be scaled up to massive datasets because of the very expensive computational cost.

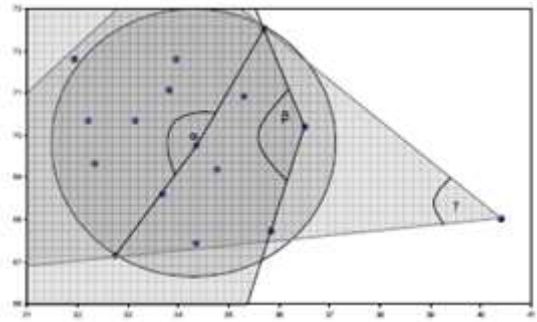

Figure 1: Intuition of angle-based outlier detection

Consider a sample data set as illustrated in fig 1. For a point within the cluster, the angles between difference vectors to pairs of other points differ widely. The variance of the angles will become smaller for points at the border of cluster. However, even here the variance is still relatively high compared to the variance of angles for real outliers. Here, the angles to most pairs of points will be small since most points are clustered in some directions.

In this paper, we observe that removing or adding an abnormal instance will cause a larger effect on principal directions than removing or adding a normal one. From this observation, we apply the "Leave One Out" (LOO) procedure to check each individual point the "with or

without" effect on the variation of principal direction. This will help us to remove the suspicious outliers in the dataset. Thus, it can be used for the data cleaning purpose. Once we have a clean dataset, we can extract the leading principal directions from it and use these directions to characterize the normal profile for the dataset. Similarly, we can evaluate the "with or without" effect of new arriving data point. That defines a suspicious score for the new arriving data point. If the score is greater than a certain threshold, we regard this point as an outlier. Based on this strategy, we proposed an on-line anomaly detection method. Intuitively, the "with or without" effect on the principle direction will be diminished for a single data point even it is an outlier when the dataset is large. To overcome this problem, we employ the "over-sampling" scheme that will amplify the "with or without" influence made by an outlier. We also are aware of computation issues in the whole process. How to compute the principle directions efficiently when the mean and covariance matrix are changed slightly is also a key issue and the tricks for matrix computation will be includes in this work as well.

2. Over-sampling Principal Component Analysis

In this section, we first introduce the classical dimension reduction method PCA briefly. The study on the influence of the variation of principal directions via LOO procedures is also be exhibited. Finally, we introduce the over-sampling scheme in PCA to emphasize the influence of an abnormal instance. In addition, an effective computation for computing the covariance matrix and estimating principal directions in LOO procedure is also proposed.

2.1 Principal Component Analysis

PCA is an unsupervised dimension reduction method. It can retain those characteristics of the data set that contribute most to its variance by keeping lower-order principal components. These few components often contain the "most important" aspects of the data. Let $A \epsilon R^{p \times n}$ be the matrix and each column, $x_i \epsilon R^p$, represents an instance. PCA involves the eigenvalue decomposition in the covariance matrix of the data. Its formulation is solving an eigenvalue problem as follows:

$$\sum_A \Gamma = \lambda \Gamma \text{ , (1)}$$

Where $\sum_A = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$ is the covariance matrix, μ is the grand mean, and the resulting Γ is the eigenvector set. In practical, some eigenvalues have little contribution to variance and can be discarded. It means that we only need to keep few components to represent the data. In addition, PCA explains variance and is sensitive to outliers. A few points distant from the center would have a large influence on variance and its principal directions. In other words, these first few principal directions will be influences seriously if our data contain some outliers.

2.1. Effects if an outlier on Principal directions

Based on the concept that we mentioned in the previous section, PCA is sensitive to outliers and we only need few principal components to represent the main data structure. That is, an outlier or a deviated instance will cause a larger effect on these principal directions, hence, we explore the variation of principal directions when removing or adding an instance. This concept is illustrated in fig. 2 where the clustered blue circles represent the normal data, the red square represents outlier, and the green arrows is the first principal direction is affected hen we remove an outlier. The first principal direction is changed and forms a larger angle between old one and itself. In this case, the first principal direction will not be affected and only form an extremely small angle between the old first principal direction and the new one if we remove a normal instance. Via this observation, we use LOO procedure to check each individual point the "with or without" effect. On the other hand, we might have the pure normal data in the hand. In this case, we use the same concept in LOO setting but with incremental strategy. i.e. adding an instance to see the variation of the principal directions. Similarly, adding a normal data point will create a smaller angle between the old one and itself while it will form a larger angle with adding an outlier (from the left panel to right panel in fig 1;). We check the variation of the principal directions is significant.

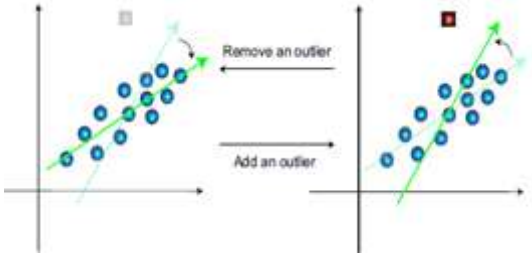

Figure 2: the illustration for the effect of an outlier on the first principal direction

In summary, we find that the principal directions will be effected with removing an outlier while the variation of the principal will be smaller with the removing a normal instance. This concept can be used for identifying the anomaly or outliers in our data. On the contrary, adding an outlier will also cause a larger influence on the principal directions while the variation of the principal directions will be smaller with adding a normal one. It means that we can use the incremental strategy to detect the new arriving abnormal data or outliers. In other words, we explore the variation of the principal directions with removing or adding a data point and use this information to identify outliers and detect new arriving deviated data.

2.2. Oversampling Principal Components Analysis

In the previous section we identify that outliers in our data and detect the new arriving outliers through the variation of the principal directions. However, the effect of "with or without" a particular data may be diminished hen the size of the data is large. On the other hand, the computation in estimating the principal directions will be heavy because we need to recompute the principal directions many times in LOO scenario.

In order to overcome the first problem, we employ "over-sampling" scheme to amplify to the outlierness on each data point. For identifying an outlier via LOO strategy, we duplicate the target instance instead of removing it. That is, we duplicate the target instance many times (10% of the whole data in our experiments) and observe how much

variation the principal directions vary. With this over-sampling scheme, the principal directions and mean of the data will only be affected slightly if the target instance is a normal data point shown in fig 3(a). On the contrary, the variations will be enlarged if we duplicate an outlier shown in fig: 3(b).

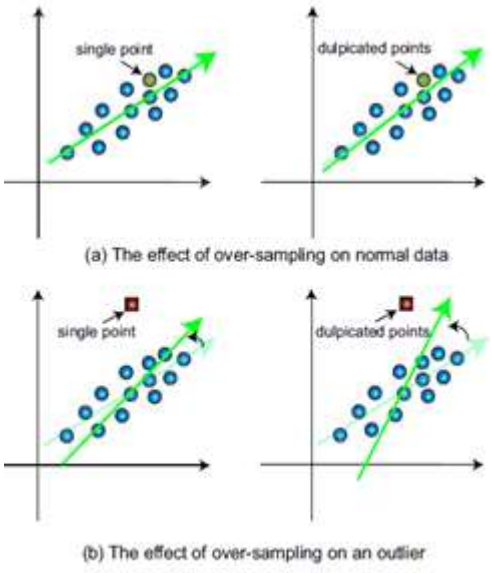

Figure 3: The effect of over-sampling on an outlier and normal instances

On the other hand, we also can apply over-sampling scheme in the LOO procedure with incremental case. The main idea is to enlarge the difference of the effect between a normal data point and an outlier. Based on the oversampling PCA, we make the idea discussed in this section is more practical.

Algorithm *1: Over-sampling principal component analysis outlier detection for data cleaning*
Input: a data matrix $A \epsilon R^{p \times n}$ *and the ratio r*
Output: the suspicious outlier ranking for their data.

1. *Compute outer-product* $Q=\frac{AA^T}{n}$, *the mean μ, and the first principal direction v*
2. *Using LOO strategy to duplicate the target instance* x_t *and compute the adjusted mean vector* $\breve{\mu}$ *and covariance matrix* $\breve{\sum}$*:*
   a. $\breve{\mu} = \frac{\mu+r.x_t}{1+r}$,
   b. $\breve{\sum}=\frac{1}{1+r}Q+\frac{r}{1+r}x_t x_t^T-\breve{\mu}\breve{\mu}^T$
3. Extract the adjusted first principal direction $\tilde{v}$ and compute the cosine similarity of v and $\tilde{v}$.
4. Repeat the step2 and 3 until scanning all the data.
5. Ranking all instances according to their suspicious outlier scores (1- | cosine similarity|).

Where $A \epsilon R^{p \times n}$ is the data matrix, x_t is the target instance and r is the parameter of the proportion of the whole data in duplicating x_t. from the equation

$$\breve{\sum}=\frac{1}{1+r}Q+\frac{r}{1+r}x_t x_t^T-\breve{\mu}\breve{\mu}^T \quad (2)$$

It shows that we can keep the matrix Q in advance and need not to recompute it completely in LOO procedure. In extracting the first principal direction, we also apply the power method for fast computation. Power method [7] is an eigenvalue algorithm for computing the greatest eigenvalue and the corresponding eigenvector. Given a matrix M, this method starts with an initial normalized vector μ_0, which could be an approximation to the dominant eigenvector or a nonzero random vector, then iteratively computes the u_{k+1} as follows:

$$u_{k+1}=\frac{Muk}{||Muk||} \quad (3)$$

The sequence $\{u_k\}$ converges on the assumption that there exists an largest eigenvalue of M in absolute value. From (3), we can see that power method does not compute matrix decomposition but only uses the matrix multiplication. Based on this property, the power method can converge rapidly and make our LOO procedure faster. On the other hand, if we want to find the remaining eigenvectors, we could use definition process [7]. Note that we only use the first principal component in our experiments so we only apply the power method in estimating the first principal detection.

3. osPCA for online Anomaly Detection

In this section we present the framework of our data analysis. There are two phases in our framework, data cleaning and online anomaly detection. In the data cleaning phase, the goal is to identify the suspicious outliers. First, we over-sample each instance with LOO strategy to see the variation of the first principal direction. Here we use the absolute value of cosine similarity to measure the difference of the first principal direction and define "one minus the absolute value of cosine similarity".

Algorithm 2: Over-sampling Principal Component Analysis for On-line Anomaly Detection

Input: The scaled outer-product matrix $Q=\frac{AA^T}{n}$, *the mean vector μand the first principal direction v of the normal data, the ratio r, and threshold h, and the new arriving instance x.*
Output: x is an outlier or not

1. compute the updated mean vector $\breve{\mu}$ and covariance matrix $\breve{\sum}$:
   i. $\breve{\mu} = \frac{\mu+r.x_t}{1+r}$,
   ii. $\breve{\sum}=\frac{1}{1+r}Q+\frac{r}{1+r}x_t x_t^T-\breve{\mu}\breve{\mu}^T$.
2. Extract the updated first principal direction $\tilde{v}$ and compute the cosine similarity of v and $\tilde{v}$.
3. Check the cosine similarity of v and $\tilde{v}$and see if it is higher the specified threshold h.

A higher suspicious outlier score implies the higher probability of being an outlier. Once we have the suspicious outlier scores for each instance, we can rank the instances and filter out the outliers in the given data according to the ranking. The over-sampling principal component analysis outlier detection algorithm (OPCAOD) for data cleaning is describing in algorithm 1.

After filtering the suspicious points, we can get the pure normal data and apply the online anomaly detection which is not suitable for LOF and ABOD. Nevertheless, the quick updating of the principal directions in our proposed method can satisfy the online detecting demand. In this phase, the goal is to identify the new arriving instance to check the variation of the principal directions. However, how to

determine the threshold for identifying an abnormal instance is a problem. In order to overcome this problem, we use some statistics to set the threshold. The idea is calculating the mean and standard deviation of the suspicious scores which are computed from all normal data points. Once we have the mean and standard deviation, a new arriving instance will be marked if its suspicious score is higher than the mean plus a specified multiple of the standard deviation. The over-sampling principal component analysis for on-line anomaly detection (OPCAAD) is also described in algorithm 2.

4. Online Anomaly Detection for Practical Scenario

Compared to the power method or other popular anomaly detection algorithms, the required computational costs and memory requirements are significantly reduced, and thus our method is especially preferable in online, streaming data, or large scale problems.

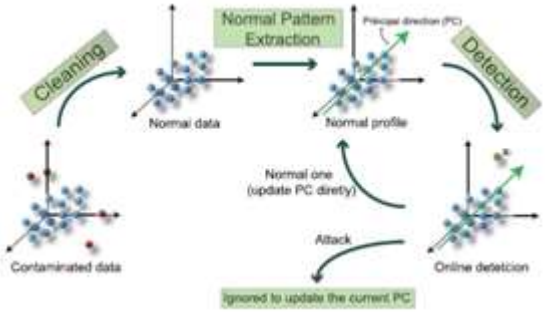

Figure 4: online anomaly detection frameworks

Online anomaly detection for practical scenario for online anomaly detection applications such as spam mail filtering, one typically designs an initial classifier using the training normal data, and this classifier is updated by the newly received normal or outlier data accordingly. However, in practical scenarios, even the training normal data collected in advance can be contaminated by noise or incorrect data labeling. In order to construct a simple yet effective model for online detection, one should disregard these potentially deviated data instances from the training set of normal data. The flowchart of our online detection procedure is shown in fig (4).

5. Conclusion and Future Work

We have explored the variation of principal directions in the leave one out (LOO) scheme. From the experimental results, we demonstrated that the variation of principal directions caused by outliers indeed can help us to detect the anomaly. We also proposed the over-sampling PCA to enlarge the outlierness of an outlier. In addition, an effective computation for computing the covariance matrix and estimating principal directions in LOO is also proposed for reducing the computational loading and satisfying the online detecting demand which is not suitable for LOF and ABOD. On the other hand, our proposed PCA based anomaly detection is suitable for the extremely unbalanced data distribution (such as network security problems). In the future, we will also study how to speed up the procedure via online learning techniques (ie., develop a quick adjusting for the principal directions directly).

6. Acknowledgment

Most of all, I shall give all glory, honor and thank to my parents. They made me as I am. Then there are a few people I would like to thank, my guide Mr. D Ratna Kishore sir, who spared no effort to ensure that I have everything I needed. Finally, my friends who gave their time, love and energy.

References

[1] Breunig, M.M., Kriegel, H.P., Ng, R., Sander, L,: LOF: Identifying density-based local outliers. In Proc. Of the 2000 ACM SIGMOD Int. Conf. on Management of Data, Dallas, Texas (2000).
[2] Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Computing Surveys (2009).
[3] Hawkins, D.: Identification of Outliers. Chapman and Hall, London (1980)
[4] Huang, L., Nguyen, X., Garofalakis, M., Jordan, M.I., Joseph, A., Taft, N.: In-network pca and anomaly detection. In: Advances in Neural Information Processing Systems, vol. 19, pp. 617–624. MIT Press, Cambridge (2007)
[5] Kriegel, H.-P., Schubert, M., Zimek, A.: Angle-based outlier detection. In: Proc. of 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Las Vegas, NV (2008)
[6] Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., Srivastava, J.: A comparative study of anomaly detection schemes in network intrusion detection. In: Proc. of the Third SIAM Conference on Data Mining (2003)
[7] G. H. Golub, C.F. Van Loan, Matrix Computations. Johns Hopkins University Press, Baltimore Md, USA, 1983.

Author Profile

Y. Srilakshmi, pursing Master's degree in computer science and engineering, JNTU KAKINADA. Her research interests include data mining, operations research, optimization, information security and anomaly detection in machine learning.

Mr. D Ratna Kishore received the masters' degree and pursing PhD degree in Computer Sciences. He is an Assistant Professor in the department of Computer Sciences. He has 3 years experience in teaching. His research interest includes operation research, machine learning, optimization and pattern recognition.