

Outlier Detection Based on Local Kernel Regression for Instance Selection

Qinmu Peng¹, Yiu-ming Cheung^{1,2 *}

¹Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China

²United International College, Beijing Normal University – Hong Kong Baptist University, Zhuhai, China

E-mail: pqinmu@gmail.com, ymc@comp.hkbu.edu.hk

Received 4 July 2012

Accepted 11 March 2013

Abstract

In this paper, we propose an outlier detection approach based on local kernel regression for instance selection. It evaluates the reconstruction error of instances by their neighbors to identify the outliers. Experiments are performed on the synthetic and real data sets to show the efficacy of the proposed approach in comparison with the existing counterparts.

Keywords: Outlier Detection; Instance Selection; Local Kernel Regression.

1. Introduction

Instance selection, which is to clean the noise-corrupted and redundant data, has been widely used in real world applications such as marketing research¹ and data mining². It is able to circumvent the over-fitting problem and improve a learner's performance, meanwhile speeding up the learning processing. Instance selection is therefore regarded as an important preprocessing step in data analysis.

Typically, the instance selection can be made via outlier removal. In general, an outlier¹⁸ is considered as an observation that deviates too much from other observations, i.e. an outlier is generated by a different mechanism compared to the other observations. The task of outlier detection is defined as follows: "Given a set of N data points or objects and the number n of expected outliers, find the top n objects that are considerably dissimilar, exceptional, or inconsistent with respect to the remaining data"^{19,20}. Thus far, various outlier detec-

tion techniques for instance selection have been proposed, which can be roughly grouped into five categories: (1) statistical or distribution-based approaches, (2) density-based approaches, (3) distance-based approaches, (4) deviation-based approaches, and (5) cluster-based approaches. Statistical-based approaches^{3,4} used a standard distribution model for data points, through which the outliers can be detected based on their relationship with the distribution model. Usually, the performance of such a method quite relies on the hypothesis model. In density-based methods, outlier is detected from the local density of observations with the different density estimation strategies. For instance, Breunig et al.⁵ proposed the local outlier factor (LOF) to score each points, depending on the local density of its neighborhood. By defining the neighborhood based on the distance to the MinPts-th nearest neighbor, Jin et al.⁶ proposed an algorithm to discover the top n outliers efficiently for a particular value of MinPts. Furthermore, Tang et al.⁷ developed the connectivity-

*Corresponding author.

based outlier factor (COF) algorithm analogous to the LOF approach. Papadimitriou et al.⁸ computed the number of neighbors for each point to evaluate the outlier-ness, in which a point whose neighborhood size significantly varies with respect to the neighborhood size of its neighbors is regarded as an outlier. In distance-based methods, it judges a point based on the distances to its neighbors. That is, normal data points have a dense neighborhood, while outliers are far apart from their neighbors. In the work^{2,9,10} by Knorr et al., a data point in the data set P is a distance-based outlier if at least a fraction β of the points in P is further than r from it. Also, Hautamaki et al.¹¹ constructed the k-NN graph for a data set, in which a vertex that has an indegree less than the user-defined threshold is an outlier. In deviation-based approaches, it groups points and considers those points as outliers that deviate considerably from the general characteristics of the groups. For example, Kriegel et al.¹² proposed the angle-based outlier detection in high dimensional space. The idea is that a point is an outlier if most of its neighbors are located in the similar directions. Otherwise, it is an inner point whose neighbors are located in different directions. Besides, some other deviation-based approaches have been proposed in work^{13,14}. In cluster-based approaches, they usually detect outliers as by-products¹⁵. Hence, these algorithms cannot properly detect the outliers under the noisy environment unless the number of clusters is known in advance. Along this line, He et al.¹⁶ proposed FindCBLOF to determine the Cluster-Based Local Outlier Factor (CBLOF) for each data point. Yang et al.¹⁷ had introduced a globally optimal exemplar-based GMM to detect the outliers, in which a Gaussian is centered at each point. The outlier factor at each point is calculated by the weighted sum of the mixture proportion with the weights representing the similarities to the other points.

In this paper, we will present an outlier detection approach based on local kernel regression (LKR) along the line of distance-based approaches. Compared to the statistical-based approaches and density-based approaches, it need not require a priori knowledge of the distribution or density function. Also, it can provide more local structure in-

formation than the deviation-based approaches, and avoid the unstableness of cluster-based approaches.

To be specific, the proposed outlier detection algorithm takes into account the local structure of the data space. We consider not only the distance between a point and its neighbors, but also the distance between neighbors. Hence, each data can be better approximated by a combination of its neighbors. Usually, the reconstruction error of an inner point is much smaller than that of a point that is on the boundary of the data distribution. Under the circumstances, we shall present an iterative method based on local kernel ridge regression to detect the outliers inward starting from the utmost outlier which has the greatest reconstruction error. Experiments have shown its promising results on synthetic and real data sets in comparison with the existing counterparts.

The remainder of this paper is organized as follows. Section 2 briefly overviews the kernel ridge regression. The details of the proposed approach, including the time complexity analysis, are given in Section 3. Section 4 shows the experimental results. Finally, we draw a conclusion in Section 5.

2. Overview of Kernel Ridge Regression

Kernel ridge regression^{21,24} is an effective approach to model nonlinear regression. Given the training data $(\mathbf{x}_i, y_i)_{i=1}^N$, where \mathbf{x}_i is an input variable and y_i is the corresponding target value, kernel ridge regression is to estimate the target value using the input variables. It can be defined as

$$g(\mathbf{x}) = \sum_{i=1}^N \alpha_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i), \quad (1)$$

where $\mathcal{K}(\cdot, \cdot)$ is a kernel function, and α_i s are the coefficients which can be estimated by minimizing the following objective function:

$$\min_{\alpha} \|\mathbf{K}\alpha - \mathbf{y}\|^2 + \gamma \alpha^T \mathbf{K}\alpha, \quad (2)$$

where $\alpha = [\alpha_1, \dots, \alpha_N]^T$, $\mathbf{y} = [y_1, \dots, y_N]^T$, γ is a small positive regularization parameter, and $\mathbf{K} = [k_{i,j}]_{N \times N}$ with $k_{i,j} = \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$. In Eq.(2), the first and second terms are called the fitness item and the

regularization item, respectively. The solution of Eq.(2) is

$$\alpha = (\mathbf{K} + \gamma \mathbf{I})^{-1} \mathbf{y}, \quad (3)$$

where \mathbf{I} is an $N \times N$ identity matrix. Therefore, $g(\mathbf{x})$ can be given as

$$g(\mathbf{x}) = \mathbf{k}_x^T (\mathbf{K} + \gamma \mathbf{I})^{-1} \mathbf{y}, \quad (4)$$

where $\mathbf{k}_x = [\mathcal{K}(\mathbf{x}, \mathbf{x}_1), \dots, \mathcal{K}(\mathbf{x}, \mathbf{x}_N)]^T$ is a vector that is formed by the value of kernel function.

3. A New Algorithm for Outlier Detection using Local Kernel Regression

It is found that the reconstruction error of an inner point is usually much smaller than that of one on the boundary. Accordingly, we measure the outlier-ness of points based on this phenomenon. That is, we examine how well the feature value of each point can be estimated by its neighbors. In this paper, we utilize a local kernel ridge regression to perform the estimation.

Given a training data \mathbf{x}_i and $\{(\mathbf{x}_j, l_{rj})\}_{\mathbf{x}_j \in N_i}$, N_i denotes the neighbors of \mathbf{x}_i , and l_{rj} is the r th feature of \mathbf{x}_j . We need to train a kernel ridge regression model to estimate the value l_{ri} , i.e. the r th feature of \mathbf{x}_i . Based on Eq.(4), we can obtain the equation to denote the local kernel ridge regression model for \mathbf{x}_i :

$$g_{N_i}(l_{ri}) = \mathbf{k}_{N_i}^T (\mathbf{K}_{N_i} + \gamma \mathbf{I})^{-1} \mathbf{l}_{rN_i}, \quad (5)$$

where $g_{N_i}(\cdot)$ is the estimation of l_{ri} as given $\{(\mathbf{x}_j, l_{rj})\}_{\mathbf{x}_j \in N_i}$, \mathbf{K}_{N_i} is the kernel matrix, i.e. $\mathbf{K}_{N_i} = [\mathcal{K}(\mathbf{x}_m, \mathbf{x}_n)]$, $\mathbf{x}_m, \mathbf{x}_n \in N_i$, and \mathbf{l}_{rN_i} denotes the vector of l_{rj} s for all $\mathbf{x}_j \in N_i$. We let

$$\beta_{N_i}^T = \mathbf{k}_{N_i}^T (\mathbf{K}_{N_i} + \gamma \mathbf{I})^{-1}. \quad (6)$$

Subsequently, Eq.(6) can be rewritten as:

$$g_{N_i}(l_{ri}) = \beta_{N_i}^T \mathbf{l}_{rN_i}. \quad (7)$$

Hence, the reconstruction error for the r th feature of \mathbf{x}_i is estimated as

$$\begin{aligned} Re(l_{ri}) &= (g_{N_i}(l_{ri}) - l_{ri})^2 \\ &= (\beta_{N_i}^T \mathbf{l}_{rN_i} - l_{ri})^2. \end{aligned} \quad (8)$$

In general, the value of each feature in an instance can be quite different, thus $Re(l_{ri})$ for each feature needs to be normalized. Let diagonal matrix D be $diag(D_{11}, \dots, D_{kk})$, and $D_{ii} = \frac{1}{|N_i|} \sum_{\mathbf{x}_j \in N_i} (\mathbf{x}_i - \mathbf{x}_j)^2$. Obviously, D_{ii} indicates the density around \mathbf{x}_i . The weighted variance $V_w(r)$ for each feature is estimated as in work²². That is,

$$V_w(r) = \sum_{i=1}^N (l_{ri} - \mu_r)^2 D_{ii}, \quad (9)$$

where $\mu_r = \sum_i l_{ri} \frac{D_{ii}}{\sum_i D_{ii}}$ is the weighted mean of the r th feature. Suppose each instance \mathbf{x}_i contains M features. The normalized total reconstruction error RE_i is given as follows:

$$\begin{aligned} RE_i &= \sum_{r=1}^M \frac{Re(l_{ri})}{V_w(r)} \\ &= \sum_{r=1}^M \frac{(g_{N_i}(l_{ri}) - l_{ri})^2}{V_w(r)}. \end{aligned} \quad (10)$$

From Eq.(8), the total reconstruction error RE_i can then be written as:

$$RE_i = \sum_{r=1}^M \frac{(\beta_{N_i}^T \mathbf{l}_{rN_i} - l_{ri})^2}{V_w(r)}. \quad (11)$$

We first select the utmost outliers on the boundary, which has the greatest reconstruction error. After removing it, the points surrounded by the outliers on the boundary then become new outliers. We therefore further remove them in the next iteration. By repeating these procedure, all outliers can be removed finally.

Since it is hard to know the ratio of outliers in advance as given a data set, a common approach is to find the top n outliers from the data sets, e.g., see the works^{19,20}. Accordingly, we propose the *algorithm LKROut* that receives an input data set X including N instances, the number n of top outliers to find, and the number k of neighbors to consider; the set *OutS* consisting of the detected outliers is the output of this algorithm.

The details of *algorithm LKRout* are summarized as follows:

Algorithm $OutS = LKROut(X, n, k)$ **begin**Construct neighbor graph \mathbf{G} and kernel matrix \mathbf{K} ;Initialize Outlier Set: $OutS = \emptyset$; $\tilde{n} = 0$; $\tilde{N} = N$;**while** ($\tilde{n} \leq n$)**do begin**Estimate $g_{N_i}(l_{ri})$ for each instance by Eq.(7);Calculate RE_i for each instance by Eq.(11);Sort all $\{RE_i\}$ in descending order;Choose the top p instances as Outliers: $OutL$;Update $OutS$: $OutS \leftarrow OutS \cup OutL$;Update $\tilde{n} = \tilde{n} + p$, $\tilde{N} = \tilde{N} - p$;Remove $OutL$ from data set;Update \mathbf{G} and \mathbf{K} ;**end;**return $OutS$;**end.**

In the above algorithm, the parameter p in each iteration can be set according to the reconstruction errors, supposing the elements in $\{RE_i\}$ are sorted in descending order. The top p instances with large reconstruction errors are chosen as outliers which account for a small portion of this data set. That is, the p is set according to the formula $p/\tilde{N} \leq \eta$, where \tilde{N} is the number of instances in the data set, η is a small ratio. Often, it just iterates several times to find the top n outliers in data instances.

The complexity analysis of key steps in the proposed algorithm is given as follows: The complexity of construction neighbor graph and kernel K is $O(N^2M)$. For local kernel regression in Eq.(7), it is $O(Nk^3)$ and the complexity for reconstruction error in Eq.(11) is $O(kNM)$. Usually, the neighbor size meets $k \ll N$. Therefore, the overall complexity is about $O(tN^2M)$, where t is the iteration times and M is the feature dimension, respectively.

In general, the proposed method is different from the nearest-neighbor methods, e.g. see papers^{2,9,10,11}. Essentially, these nearest-neighbor methods consider the distance between point \mathbf{x}_i and its nearest-neighbor only without taking into account the distance between the neighbors. Actually, the distance between the neighbors can provide more relationship information. Accordingly, the proposed method considers not only the distance between point and its neighbors, but also the distance be-

tween the pairs of neighbors by weighing the influence of the relevant neighborhood point²³. Since such a way would be better to reveal the local relationship of data, it is therefore expected that the proposed method generally outperforms these nearest-neighbor methods.

4. Experimental Results

The experiments were performed on several synthetic and real-world data sets to show the performance of the proposed algorithm in comparison with the existing counterparts. In general, local outlier detection approaches have shown better performance than global models. Accordingly, we chose two of the most popular local approaches, namely the LOF⁵ and angle-based outlier detection (ABOD)¹², as counterparts.

In all experiments, we utilized the Gaussian kernel, i.e. $\mathcal{K}(\mathbf{x}_m, \mathbf{x}_n) = \exp(\|\mathbf{x}_m - \mathbf{x}_n\|^2/h)$, and the parameter h and the ridge parameter γ were set at 100 and 0.1, respectively. To measure the performance of each algorithm, we utilized the Detection rate (True positive rate), the False alarm rate (False positive rate) from the work¹⁷, and the ROC curves, which are defined as follows:

$$\begin{aligned} \text{Detection Rate} &= TP/(TP + FN) \\ \text{False alarm Rate} &= FP/(FP + TN). \end{aligned} \quad (12)$$

As shown in Table 1, the detection rate reveals the relative ratio of correctly identified outliers, while the false alarm rate is the percentage of normal data records misclassified as outliers. Furthermore, the ROC curve is defined by False positive rate and True positive rate as x- and y-axes, respectively, which depicts the relative trade-off between Detection rate (benefits) and False alarm rate (costs).

Table 1. Performance indicator.

	Predicted Outliers	Predicted Normal
Actual Outliers	TP (True positive)	FN (False negative)
Actual Normal	FP (False positive)	TN (True negative)

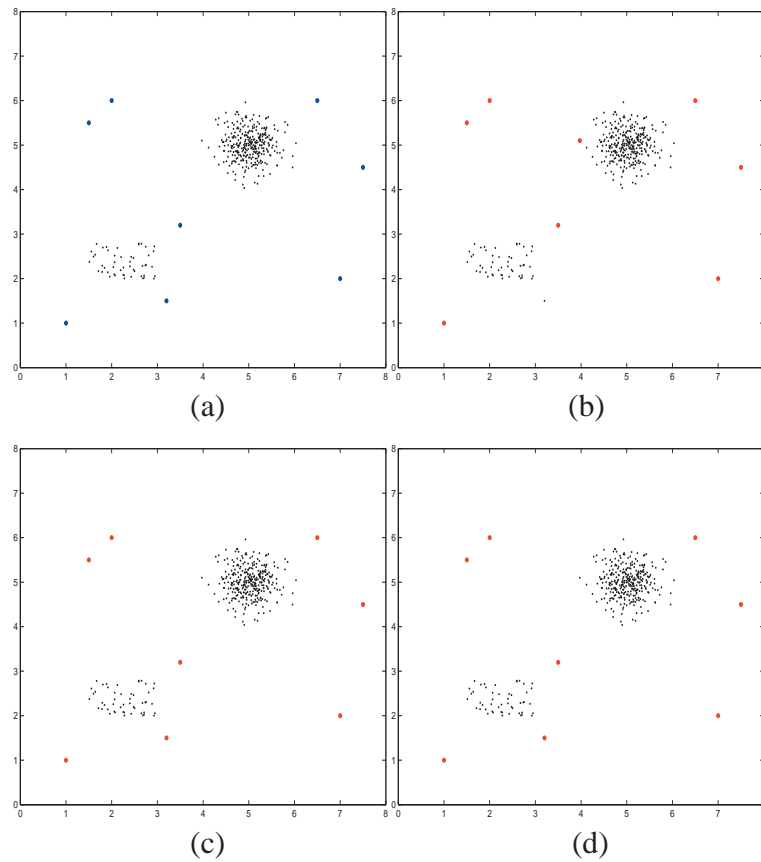


Fig. 1. Outlier detection results. (a) Data points: normal points are small dots in black, while the outliers are large dots in blue. (b) LOF method; (c) ABOD method; (d) Proposed method. Note that the large dots in red are the outliers detected by these methods.

4.1. Synthetic data

4.1.1. Synthetic data with several noisy points

The synthetic data were utilized to evaluate the performance of three approaches. In Fig. 1 (a), there are eight outliers (marked with blue dots) and two clusters with different densities, 50 points are in the sparse cluster and 318 points are in the dense cluster. From Fig. 1 (b-d), it can be observed that LOF method based on local neighbors is unable to detect all outliers, while ABOD and the proposed method are able to detect all of them in this environment. A plausible reason is that the LOF just considers the local distance between the point and its neighbors, but ignores the distance between the pairs of neighbors. As a result, the LOF, as well as the other neighbor-

based methods, is difficult to detect the outliers with the varied densities. In contrast, the proposed approach considers both distances which provide more information about the local domain. As for the ABOD approach, it is angle-based one that describes the divergence of points with respect to one another. Hence, unlike the distance-based approach, it can perform well in this case.

4.1.2. Synthetic data with many noisy points

In this experiment, more noisy points were added into the data set to evaluate the capability of the three approaches.

750 normal data points were randomly generated from a homogeneous distribution model, which

formed the shape “H”. Also, the extra 150 data points were generated by a random distribution that was independent of the model of the normal data. The synthetic data set is shown in Fig. 3 (a). We compared the proposed algorithm with the ABOD and LOF approaches on this 2-D data set. In these methods, the neighborhood size was set at 10. In order to evaluate the capability of each method to identify the most likely outliers, we utilized the true positive rate and false positive rate, and drew ROC curves. That is, we successively retrieved the most likely outliers until all outliers were retrieved. The results of the proposed approach are shown by ROC curves in Fig. 2 in comparison with the other two approaches. It can be seen that the proposed approach outperforms the other methods, and the LOF method has poor performance on this data set.

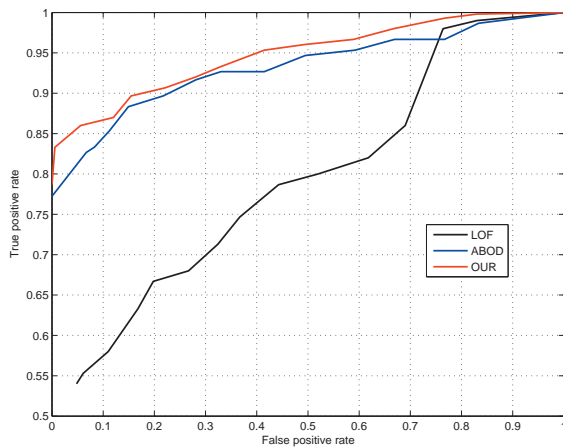


Fig. 2. ROC curves on the synthetic data.

In addition, we showed a snapshot of results by retrieving the top 15% of data points as outliers. From Fig. 3 (b-d), it can be seen that most outliers have been successfully detected by the ABOD and the proposed approach. Nevertheless, compared to the proposed method, some obvious outliers (marked in black box) were undetectable by the ABOD method. Also, the LOF approach was not suitable to deal with them even with the different neighborhood sizes.

[†]UCI ML Repository: <http://archive.ics.uci.edu/ml/datasets.html>

4.2. Real data

We further demonstrated the effectiveness of the proposed method on two real-world datasets.

4.2.1. Breast Cancer Wisconsin Data

In this experiment, we utilized the Breast Cancer Wisconsin (original) Data Set [†] which is a medical dataset and has been used for breast tumor diagnosis. The dataset contains 11-attributed (ID, class, 9 real-valued features) 699 medical diagnosis records, in which the total valid records are 683, while the remaining 16 records have missing values. The class type is binary, i.e. “Benign” and “Malignant”. We treated the 444 records labeled “Benign” as normal data. Furthermore, a certain number of “Malignant” records were added into normal data as outliers. In this dataset, our task was to identify the outliers which were the abnormal data records, and evaluated the three different approaches.

We measured the percentage of detected outliers in top-*num* potential outliers as the detection rate, i.e. $\text{detection rate} = n_{\text{detected outliers}} / \text{num}$, where *num* is the number of outliers. Experiments were performed three times with *num* = 20. At each time, we performed ten independent trials on the dataset (normal data with randomly selected outliers) to calculate the average detection rate and standard deviation. The experimental results are shown in Table 2.

Moreover, we examined the relationship between the neighborhood size and detection rate. In this experiment, we utilized 20 outliers, which were randomly extracted from the “Malignant” records. Once again, top-20 was utilized here. Fig. 4 shows the outlier detection rate over different neighborhood sizes. It can be seen that the detection rate of the proposed method is more robust against the neighborhood size. In contrast, the LOF approach suffers from the effect of the neighborhood size.

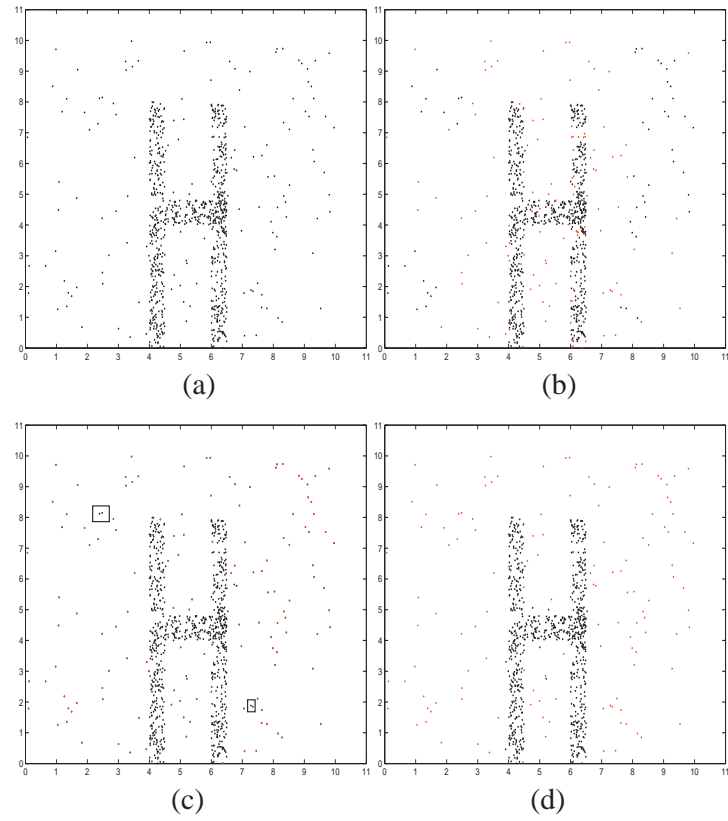


Fig. 3. Outlier detection results. (a) Data points; (b) LOF method; (c) ABOD method; (d) Proposed method. The red points are outliers detected by these methods.

Table 2. The detection rate of each method.

Number of Detected Outliers	ABOD Detection Rate	LOF Detection Rate	Our Detection Rate
20	0.6750 ± 0.0024	0.0450 ± 0.0014	0.7050 ± 0.0039
15	0.7600 ± 0.0320	0.0500 ± 0.0080	0.7800 ± 0.0200
10	0.8600 ± 0.0138	0.0620 ± 0.0093	0.8800 ± 0.0173

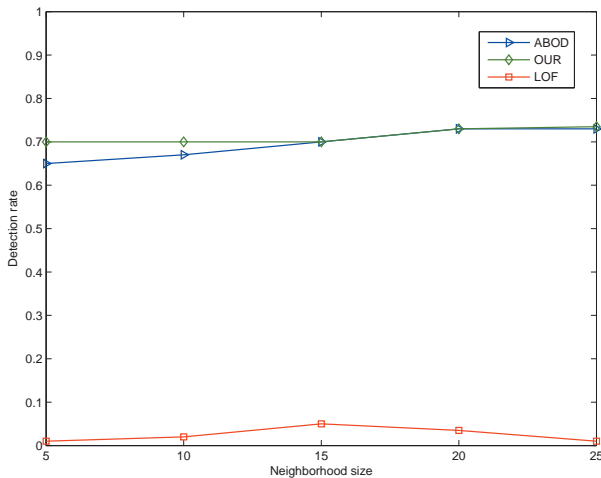


Fig. 4. Outlier detection rate over different neighborhood size by ten independent runs.

4.2.2. MPEG-7 Shape Data

We also evaluated the proposed method on the widely used MPEG-7 shape data set for detection of unusual shapes. MPEG-7 contains 70 classes with 20 shapes in each class, which are binary images. In Fig. 5, there are 20 bell shapes from one class and 4 unusual shapes from the other different classes. Hence, we should retrieve the 4 unusual shapes for the purpose of outlier detection. The shape image is represented by inner distance shape context (ID-SC) feature²⁵, which provides good shape descriptors for binary image shape. The results are demonstrated in Fig. 6. The ABOD and LOF approaches retrieve the same results that consist of three true unusual shapes. Compared to the two approaches, the proposed approach can identify all of them.

5. Conclusion

In this paper, we have proposed a new way to measure the outlier-ness of instances based on the reconstruction error of instances by their neighbors. It is found that the reconstruction error of a normal instance is usually much smaller than that of one on the boundary. Accordingly, we have presented an iterative algorithm to detect the outliers. Experimental results have demonstrated its efficacy in comparison with the existing counterparts.

Acknowledgment

The works described in this paper were supported by the Faculty Research Grant of Hong Kong Baptist University with the Project Code: FRG2/11-12/067, FRG2/12-13/082, and the NSFC under grant 61272366.

References

1. H. Nare, D. Maposa, and M. Lesaoana, A method for detection and correction of outliers in time series data, *African Journal of Business Management*, **6** (22), 6631–6639 (2012).
2. E.M. Knorr, R.T. Ng, and V. Tucakov, Distance-based outliers: algorithms and applications, *The International Journal on Very Large Data Bases*, **8**, 237-253 (2000).
3. N. Billor, A. Hadi and P. Velleman, BACON: blocked adaptive computationally efficient outlier nominators, *Computational Statistics and Data Analysis*, **34**, 279-298 (2000).
4. E. Eskin, Anomaly detection over noisy data using learned probability distributions, *Proceedings of International Conference on Machine Learning*, 255-262 (2000).
5. M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, LOF: identifying density based local outliers, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 93-104 (2000).
6. W. Jin, A. Tung, and J. H. Mining, Mining top-n local outliers in large databases, *Proceedings of the Knowledge Discovery and Data Mining Conference*, 293-298 (2001).
7. J. Tang, Z. Chen, A. Fu, and D. Cheung, Enhancing effectiveness of outlier detections for low density patterns, *Proceedings of the Pacific-Asia Knowledge Discovery and Data Mining Conference*, 535-548 (2002).
8. S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos, LOCI: fast outlier detection using the local correlation integral, *Proceedings of the International Conference on Data Engineering*, 315-326 (2003).
9. E.M. Knorr and R.T. Ng, Algorithms for mining distance-based outliers in large datasets, *Proceedings of International Conference on Very Large Data Bases*, 392-403 (1998).
10. E.M. Knorr and R.T. Ng, Finding intentional knowledge of distance-based outliers, *Proceedings of International Conference on Very Large Data Bases*, 211-222 (1999).
11. V. Hautamaki, I. Karkkainen, and P. Franti, Outlier detection using k-nearest neighbour graph, *Proceedings*

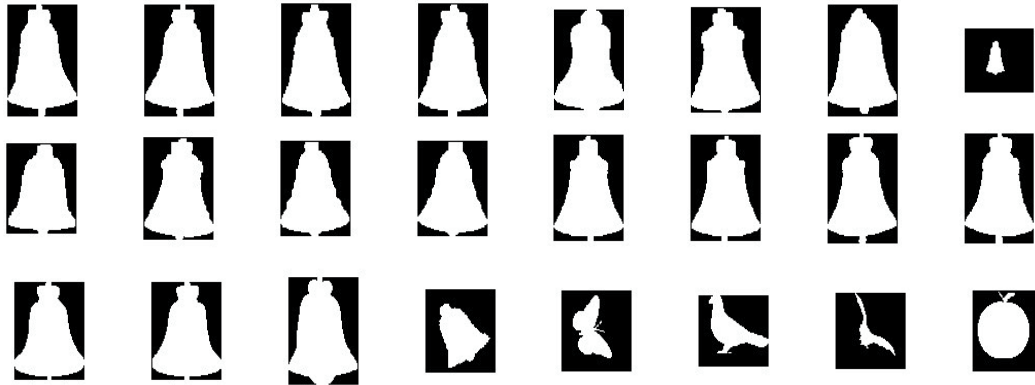


Fig. 5. The shape database for unusual shape detection.

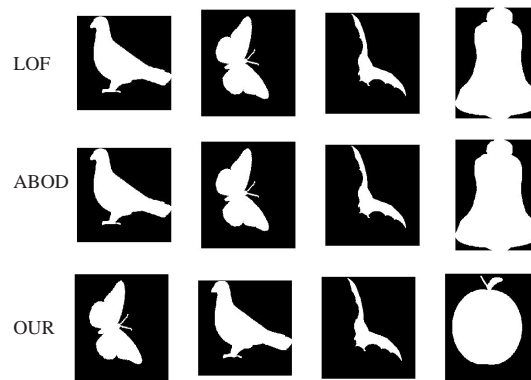


Fig. 6. Top 4 ranked unusual shapes (outliers) by the three approaches.

- of *IEEE International Conference on Pattern Recognition*, 430-433 (2004).
12. H.-P. Kriegel, M. Schubert, and A. Zimek, Angle-based outlier detection in high-dimensional data, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 444-452 (2008).
13. A. Arning, R. Agrawal, and P. Raghavan, A linear method for deviation detection in large databases, *Proceedings of Knowledge Discovery and Data Mining Conference*, 164-169 (1996).
14. S. Sarawagi, R. Agrawal, and N. Megiddo, Discovery-driven exploration of OLAP data cubes, *Proceedings of International Conference on Extending Database Technology - Advances in Database Technology*, 168-182 (1998).
15. A.K. Jain, M.N. Murty, and P.J. Flynn, Data clustering: a review, *ACM Computing Surveys*, **31** (3):264-323 (1999).
16. Z. He, X. Xu and S. Deng, Discovering cluster-based local outliers, *Pattern Recognition Letters*, **24**,1641-1650 (2003).
17. X. Yang, L. J. Latechi, and D. Pokrajac, Outlier detection with globally optimal exemplar-based GMM, *Proceedings of the SIAM International Conference on Data Mining*, 145-154 (2009).
18. D. M. Hawkins, Identification of outliers, Chapman and Hall, London (1980).
19. J. Han and M. Kamber, Data mining, concepts and techniques, San Francisco: Morgan Kaufmann, (2001).
20. F. Angiulli and C. Pizzuti, Outlier mining in large high-dimensional data sets, *IEEE Transactions on Knowledge and Data Engineering*, **17** (2), 203-215 (2005).
21. Y.M. Cheung and H. Zeng, Local kernel regression score for selecting features of high-dimensional data, *IEEE Transactions on Knowledge and Data Engineer-*

- ing, **21** (12), 1798-1802 (2009).
22. X. He, D. Cai, and P. Niyogi, Laplacian score for feature selection, *Proceedings of the Advances in Neural Information Processing Systems Conference*, 507-514 (2005).
 23. Q.M. Peng and Y.M. Cheung, Sample outlier detection based on local kernel regression, *Proceedings of IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'2012)*, 664-668 (2012).
 24. J. Shawe-Taylor and N. Cristianini, Kernel methods for pattern analysis, Cambridge Univ. Press (2004).
 25. H.B. Ling and D. W. Jacobs, Shape classification using the inner-distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29** (2): 286-299 (2007).