British Journal of Mathematics & Computer Science
4(19): 2799-2810, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

An Alternative Artificial Intelligence Technique for Detecting Outliers

Hegazy Zaher[1], Abd El-Fattah Kandil[2] and Rehab Shehata[2*]

[1]*Institute of Statistical Studies and Research Cairo University, Egypt.*
[2]*Department of Statistics, Mathematics and Insurance College of Commerce, Benha University, Egypt.*

Original Research Article

Received: 02 May 2014
Accepted: 08 July 2014
Published: 23 July 2014

Abstract

Data are rarely perfect. Whether the problem is data entry errors or rare events. Outliers have two opposing properties. They can be noises that disturb regression and classification task. On the other hand, they can provide valuable information about rare phenomena, which can lead to knowledge discovery. This paper proposes a hybrid algorithm including K Nearest Neighbor and Support Vector Machine (KSVM) that detects outliers by taking the advantages of the two intelligent techniques, Support Vector Machine (SVM) and K Nearest Neighbour (KNN). Also a global efficiency measure introduced to compare different methods. Finally, a comparison between KNN, SVM, and KSVM is conducted using detection rate, accuracy rate, false alarm rate, true negative rate and the proposed global efficiency measure based on benchmark data called Milk data.

Keywords: Outliers, outlier detection, KNN, SVM, and global efficiency.

1 Introduction

The problem of the existence of outliers in data is an important problem that has been investigated within diverse knowledge disciplines and application domains. The importance of outlier detection is due to the fact that outliers in data are translated to significant (and often critical) information in a wide variety of application domains. Therefore, outlier detection refers to the problem of finding patterns in data that do not conform to expected normal behavior. Many terms have been used including novelty detection, anomaly detection, noise detection, deviation detection, or exception mining, but all involve a similar process or goal. In addition to many different names, there are many more algorithms for solving them. *According to Barnett and Lewis* [1].

An outlier is an observation (or subset of observations) which appear to be inconsistent with the remainder of the data set.

Corresponding author: shehatarehab@yahoo.com;

1.1 Source of Outliers

1. Error: An error refers to a noise-related measurement coming from a faulty sensor. Outliers caused by errors may occur frequently, while outliers caused by events tend to have extremely smaller probability of occurrence.
2. Event: An event is defined as a particular phenomenon that changes the real-world state, e.g., forest fire, air pollution, etc. This kind of outliers usually lasts for a relatively long period of time and changes historical pattern of sensor data, [2].

There are many methods for detecting outliers. The commonly used methods of outlier detection are:

I. **Classical techniques include**:

   a. Statistical approach
   b. Clustering approach
   c. Density approach
   d. Depth approach

II. **Intelligent techniques include**:

   a. Classification approach
   b. Distance approach
   c. Information theory approach
   d. Spectral decomposition approach
   e. Visualization approach
   f. Wavelet approach

Most of the classical techniques deal with univariate variable, but most of the problems statistician faced, the data are multivariate or multi-dimension.

The rest of the article is organized as follows. Section 2 presents the evaluation methods. Section 3 involves the new technique (KSVM), its algorithm; in addition, the results of applying the mentioned methods on benchmark data called Milk data.

2 Evaluation Methods

Outlier detection algorithms are typically evaluated using the detection rate, the false alarm rate, and the accuracy rate. In order to define these metrics, let us look at a confusion matrix, shown in the following Table. In the outlier detection problem, assuming class "C" as the outlier or the rare class of the interest, and "NC" as a normal (majority) class, there are four possible outcomes when detecting outliers (class "C") - namely true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN). Global Efficiency is a general measure combining three efficiency measures of a given method (accuracy, detection rate, negative true rate), into only one value representing the total efficiency of that method. Global efficiency facilitates the comparison between different methods. It takes values between 0 and 1. More closer to 1, it indicates more efficiency.

Table 1. Confusion matrix [3]

Test Outcome	Predicted outliers class C	Predicted normal class NC
Actual outliers class C	True Positive (TP)	False Negative (FN)
Actual normal class NC	False Positive (FP)	True Negative (TN)

From the previous Table, detection rate, false alarm rate, and accuracy may be defined as follows, [3]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Detection\ Rate = \frac{TP}{TP + FN}$$

$$False\ Alarm\ Rate = \frac{FP}{FP + TN}$$

$$True\ Negative\ Rate\ (Specificity) = 1 - False\ Alarm\ rate = \frac{TN}{FP + TN}$$

$$Global\ efficiency = \sqrt[3]{Detection\ Rate * Accuracy * True\ Negative\ Rate}$$

Detection rate gives information about the relative number of correctly identified outliers, while the false alarm rate reports the number of outliers misclassified as normal data records (class NC).

We consider data from Daudin et al. [4] which consists of 8 measurements on the composition of 86 milk container samples. This data set has been studied in many papers in the outlier detection context (e.g. Atkinson, [5]), starting from Caussinus and Ruiz-Gazen [6].

Daudin et al. [4] used the data to illustrate a bootstrap study of the stability of principal components analysis; they found only two large components. But, outliers will be in the original eight-dimensional space, rather than in this reduced space. Caussinus and Ruiz-Gazen [6] extended this analysis using a generalized principal components analysis in which the covariance matrix is estimated robustly; they identified numerous outliers.

These data contains outliers. The 10 most extreme outliers-in order- observations 2, 41, 1, 44, 12, 14, 74, 13, 15 and 3 were all identified outlying by all 100 forward searches, (Atkinson [5]).

A classical multivariate analysis using the sample mean and covariance estimate would not detect anything. So, KSVM is used to detect outliers. Especially, it is an intelligent technique.

Let us consider a data set D with p features and n instances. In a supervised classification context, we must know also the classes where each of the instances belongs. It is quite common involve the classes as the last column of the data matrix. The goal is to find out all the instances that seem to be remarkable these will be the multivariate outliers. By using SPSS software program, we test the data (Milk data) by boxplot test for univariate data. One might think that multivariate outliers can be detected based on the univariate outliers on each feature but as it is shown in the Fig. 1 this is not true. The instance appearing in variables 1, 2, 7 and 8 (which is sample number 12) is an outlier but it is false outlier in each characteristic. Additionally, an instance can be considered outliers in several features but the total instance might be not a multivariate outlier. There are various methods to detect multivariate outlier.

For example, we want to find out the multivariate outliers in each of the variables of the dataset milk by building box plots. People in the data mining community prefer to rank the instances using multi-criteria measures rather to classify the instances outliers and non-outliers. Rocke and Woodruff [7] stated the Mahalanobis distance works well in identifying diversified outliers, Mahalanobis [8]. However, in data with clustered outliers the Mahalanobis distance measures do not give good results for detecting outliers. The masking and swamping effects will affect Data sets with multiple outliers or clusters of outliers.

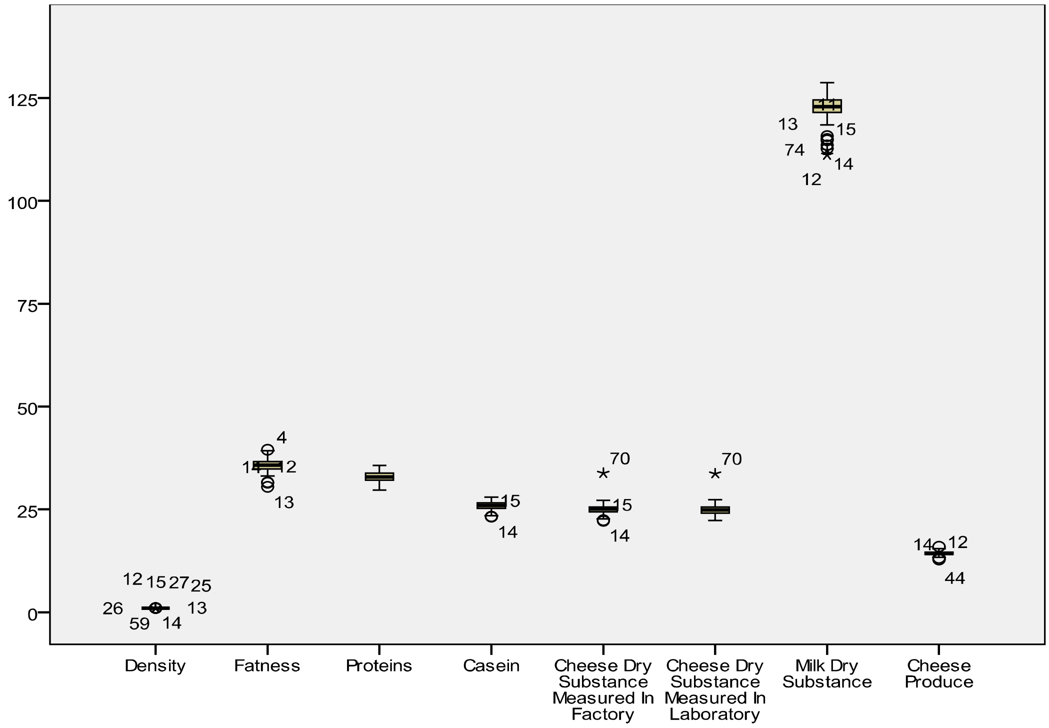

Fig. 1. Detecting univariate outliers by box plots in the Milk dataset

- ***Masking effect***. It is said that an outlier masks a second one close by if the latter can be considered outlier by himself, but no anymore if it is considered along with the first one. Equivalently, after the deletion of one outlier, another instance may appear as an outlier. Masking happens when a group of outliers skews the mean and covariance estimations toward it, and the obtained distance of the outlier from the mean is small.
- ***Swamping effect***. An outlier swamps another one if the latter can be viewed outlier just under the presence of the first one In other words after the deletion of one outlier other outlier may become a “good” instance. Swamping occurs when a group of outlying instances skews the mean and covariance estimates toward it and away from other

"good" instances, and the resulting distance from these "good" points to the mean is large making them look like outliers.

3 KSVM Methods

The KNN method is excellent with its simple training and classifying algorithms, according to the fact that KNN implements instant-based learning algorithm. Hence, it does not need construction of a feature space until it receives a command to classify an input sample. Besides this, it can manage multiclass dataset sufficiently by requiring a few training samples. Most distance metrics, as that of nearest neighbors, are supposed to become less sensible as the dimensionality of the noisy data increases, thus limiting the performance of NN when applied to data classification. In other words, the performance of KNN classifier is severely degraded by noises. In many cases, KNN is often used as the preliminary classifier to prune the training samples and eliminate those samples effectively to reduce the sample size for better classification, Knorr [9].

Support Vector Machines (SVMs) were originated from statistical learning theory. The concept of SVMs was introduced by V. N. Vapnik in the late 1970's. SVMs are now regarded as an important example of "kernel methods", arguably the hottest area in machine learning. Machine learning algorithm is the ability of a computer algorithm to recognize patterns that have occurred repeatedly and to improve its performance based on past experiences. SVMs separate different classes by finding optimal classification hyper planes from training, leading to better generalization capability compared with other methods. In SVMs learning, kernel methods are often used to map the data vectors in the input space into a higher dimension feature space thereby converting linearly non-separable case into a separable case. The construction of a linear classification hyper plane in this high dimension feature space is equivalent to a nonlinear decision hyper plane in the input lower dimension space, Vapnik [10].

Support Vector Machine is a learning algorithm based on the principle of margin maximization. Comparative studies have found SVMs to perform as well as or better than most prevalent learning methods.

Rong et al. [11] were the first who proposed KSVM classifier as an improvement to the accuracy of SVM classifier.

Then, Shen and Lin [12] presented a new binary classifier that combines SVM with KNN method for gene expression data classification to enhance disease diagnosis. Regular SVM classification yields lower classification accuracy whenever the sample data of the two classes in a binary classification are all near to the separating hyper plane. When traditional KNN classification has been applied to achieve gene expression data classification, KNN distance functions are expected to have lower classification performance as the dimensionality of the noisy data increases. The KNN and SVM hybrid classification system introduced in this work, coined as KSVM, is explained due to the reality that traditional SVM is not able to handle the situation when the training samples are close to the optimal separating hyper plane Shen and Lin [12].

Zhang et al. [13] suggested an integration of nearest neighbor classifier and support vector machine to classify visible objects, coined as KSVM. The suggested KSVM hybrid allocation model can be used for large and multiclass datasets with reasonably lower computational complexity in both training and classifying stage. This property of lower computational complexity is inherited from KNN classification approach, which does not need construction of a feature space, and it can handle and manage multiclass datasets with high dimensionality. Due to this fact, the proposed KSVM model is able to provide a more flexible framework from a computer vision perspective as the emphasis is on similarity of objects.

Mei et al. [14] proposed a hybridized algorithm that uses both (KNN) and (SVM) classifiers, known as KSVM, to examine gene expression for cancer classification, in both binary and multi-class categorization. They stated that accurate diagnosis and classification is the last goal for successful treatment of illness.

KNN technique has used by KSVM hybrid classification method as the beginning process of its classification task. In the state that there is no opposition in the classification of a given input data, where all k neighbours are from a same class, a direct classification task is performed by the simple NN algorithm without the involvement of SVM approach. On the other hand, if the k neighbours are not in the same class, the hybrid classification procedure is used to achieve the classification task.

In the proposed KSVM hybrid classification method, Nearest Neighbor (NN) algorithm is used to prune the original training set, in that case SVM classification of this hybrid model does not require to train with the original training set, but with a pruned training set. This guarantees a faster training process for SVM with a smaller training set, without losing the sufficiency of classification performance. For the pruned training set, pair-wise distances between the training samples are calculated and the distance matrix will be converted to kernel matrix as the training data for SVM. The general idea of this technique is to train the SVM until it preserves the distance function on the set of neighbors and to find close neighbors to an input sample.

SVM is sensitive to these samples intermixed in another class or these boundary samples. The existence of samples in the overlapped region may be harmful to the performance of SVM. The hybridized classifier can improve the precise of classification, so hybridized of KNN classifiers and SVM is proposed to improve the performance of classification.

The first KNN is used to prune training samples and the second KNN is combined with SVM to classify the Milk samples (cases). We first yield the distances matrix which is a symmetrical matrix containing the Euclidean distance between each pair of samples. Then the k nearest neighbors for each sample is sought. In the first KNN, if the class label of training sample is same as the label of the majority of its k nearest neighbors, the training sample is reserved, whereas others are pruned. For the pruned samples set, the second KNN and SVM are applied to classify. If k nearest neighbors has all the same labels, the sample is labeled. Otherwise, SVM will be applied to classify the rest sample. The SVM-KNN steps are described as follows.

Step 1: Compute distances of the test to all training examples and pick the nearest K neighbors;
Step 2: If the K neighbors have all the same labels, the test is labeled and exit; else, compute the pairwise distances between the K neighbors;
Step 3: Convert the distance matrix to a kernel matrix and apply SVM;
Step 4: Use the resulting classifier to label the test.

To implement multidimensional SVM in step 3, three variants from the statistics and learning literature have been tried on small number of samples from our data sets. They produce roughly the same quality of classifiers and the SVM is chosen for its better speed.

The new algorithm of KSVM is slow mainly because it has to compute the distances of the test to all training examples. In our setting, this translates into the practice of computing a "crude" distance to prune the list of neighbors before the more costly "accurate" distance computation. The reason is simply that if the crude distance is big enough then it is almost certain that the accurate distance will not be small. This idea works well in the sense that the performance of the classifier is often unaffected whereas the computation is orders-of-magnitude faster.

An additional trick to speed up the algorithm is to cache the pairwise distance matrix in step 2. This follows from the observation that those training examples who participate in the SVM classification lie closely to the decision boundary and are likely to be invoked repeatedly during test time.

So far, there are two perspectives to look at KSVM: it can be viewed as an improvement over NN classifier, or it can be viewed as a model of the discriminative process. From a machine learning perspective, it can also be viewed as a continuum between KNN and SVM: when K is small, the algorithm behaves like a straightforward KNN classifiers. To the other extreme, when K = n our method reduces to an overall SVM.

4 Numerical Example on KSVM

For example, we have a telecommunications company that counts the number of calls conducted by some companies and duration of these calls within 21 days. The company classified these calls as is not outlier value and outlier values.Telecommunications company wants to test a new sample (in which number of calls is 12 and duration of these calls is 115), if it is outlier or not outlier by using KSVM method.

Day	**X1= No. of calls**	**X2= Duration of calls**	**Y= classification**
1	13	95	1=Not outlier
2	10	83	1=Not outlier
3	9	91	1-Not outlier
4	13	102	1=Not outlier
5	11	100	1=Not outlier
6	8	104	1=Not outlier
7	7	113	1=Not outlier
8	10	83	1=Not outlier
9	11	84	1=Not outlier
10	10	100	1=Not outlier
11	12	105	1=Not outlier
12	11	86	1=Not outlier
13	10	100	1=Not outlier
14	23	111	-1=outlier
15	22	119	-1=outlier
16	23	122	-1=outlier
17	25	115	-1=outlier
18	29	106	-1=outlier
19	33	102	-1=outlier
20	35	111	-1=outlier
21	24	121	-1=outlier

To apply SVM-KNN method, we have to do the following steps:

Step1: Compute distances of the test to all training examples and pick the nearest K neighbors;

In this step, we apply KNN method by using K=8. The following results are shown:

Day	**X1= No. of call**	**X2= duration of call**	**Classification**	**Squared distance (x1-12)2 +(x2-115)2**	**Rank**	**KNN K=8**	**New classification**
1	13	95	Not outlier	401	14	No	-
2	10	83	Not outlier	1028	20	No	-
3	9	91	Not outlier	585	17	No	-
4	13	102	Not outlier	170	7	Yes	Not outlier
5	11	100	Not outlier	226	10	No	-
6	8	104	Not outlier	137	5	Yes	Not outlier
7	7	113	Not outlier	29	1	Yes	Not outlier
8	10	83	Not outlier	1028	21	No	-
9	11	84	Not outlier	962	19	No	-
10	10	100	Not outlier	229	11	No	-
11	12	105	Not outlier	100	2	Yes	Not outlier
12	11	86	Not outlier	842	18	No	-
13	10	100	Not outlier	229	12	No	-
14	23	111	outlier	137	4	Yes	outlier
15	22	119	outlier	116	3	Yes	outlier
16	23	122	outlier	170	8	Yes	outlier
17	25	115	outlier	169	6	Yes	outlier
18	29	106	outlier	370	13	No	-
19	33	102	outlier	610	16	No	-
20	35	111	outlier	545	15	No	-
21	24	121	outlier	180	9	No	-
Tested sample	12	115					not labeled

Step 2: If the K neighbors have all the same labels, the test is labeled and exit; else, compute the pairwise distances between the K neighbors;

We found that the tested sample not labeled because, that there are 4 samples classified the testing sample as not outlier and 4 samples classified the testing sample as outlier. There is a condition in KSVM stated that, if the K neighbours are not in the same class (outliers or not outliers), SVM method is used to perform the classification task.

Step3: Apply SVM method by using Microsoft Office Excel. The following results are shown.

	A	B	C	D	E	F	G	H	I	J	K	L
2			w1	w2	b		objective	0.012738854				
3			-0.14013	-0.076433121	10.70700637							
4		index	x1	x2	y		w1.x1	w2.x2	constraint		score	classification
5		1	13	95	1		-1.82166	-7.261146497	1.624204		1.624204	1
6		2	10	83	1		-1.40127	-6.343949045	2.961783		2.961783	1
7		3	9	91	1		-1.26115	-6.955414013	2.490446		2.490446	1
8		4	13	102	1		-1.82166	-7.796178344	1.089172		1.089172	1
9		5	11	100	1		-1.5414	-7.643312102	1.522293		1.522293	1
10		6	8	104	1		-1.12102	-7.949044586	1.636943		1.636943	1
11		7	7	113	1		-0.98089	-8.636942675	1.089172		1.089172	1
12		8	10	83	1		-1.40127	-6.343949045	2.961783		2.961783	1
13		9	11	84	1		-1.5414	-6.420382166	2.745223		2.745223	1
14		10	10	100	1		-1.40127	-7.643312102	1.66242		1.66242	1
15		11	12	105	1		-1.68153	-8.025477707	1		1	1
16		12	11	86	1		-1.5414	-6.573248408	2.592357		2.592357	1
17		13	10	100	1		-1.40127	-7.643312102	1.66242		1.66242	1
18		14	23	111	-1		-3.22293	-8.484076433	1		-1	-1
19		15	22	119	-1		-3.0828	-9.095541401	1.471338		-1.47134	-1
20		16	23	122	-1		-3.22293	-9.324840764	1.840764		-1.84076	-1
21		17	25	115	-1		-3.50318	-8.789808917	1.585987		-1.58599	-1
22		18	29	106	-1		-4.06369	-8.101910828	1.458599		-1.4586	-1
23		19	33	102	-1		-4.6242	-7.796178344	1.713376		-1.71338	-1
24		20	35	111	-1		-4.90446	-8.484076433	2.681529		-2.68153	-1
25		21	24	121	-1		-3.36306	-9.248407643	1.904459		-1.90446	-1
26		test samp	12	115			-1.68153	-8.789808917	0		0.235669	1

Step 4: Use the resulting classifier to label the test.

As shown in the last row, the tested sample classified as 1, which indicates to be "Not outlier". This indicates that KSVM can classify outlier where KNN and SVM only cannot classify the outlier observation as outlier.

5 Application and Results

Applying mentioned illustrated methods (KNN, SVM, and KSVM) on Milk data using Matlab software, the results were shown in Table 2. As mentioned before, the data have 8 dimensions and 86 samples, 10 of them are outlier samples. In the training phase 10 normal samples (sample number 4,5,6,7,8,9,10,11,16 and 17) and 5 outlier samples (sample number 1,2,3,12 and 13) are considered. In testing phase, 66 normal samples (sample number from 18 to 40, 42, 43, from45 to 73 and from 75 to 86) and 5 outlier samples (sample number 14,15,41,44 and 74) are tested by using the three methods mentioned before. The main results are as follows:

1- The highest global efficiency score, (0.9148), for detecting outliers is satisfied by the two methods SVM (RBF) and KSVM.
2- Although the proposed method KSVM has the same efficiency as SVM (RBF) for the given data. It may be more suitable for other data because KSVM has the advantages of the two methods SVM (RBF) and KNN.

Finally the authors are considering other hybrid between different methods to find the most suitable algorithm having the highest global efficiency.

Table 2. Table of results

Evaluation method	**SVM linear**	**SVM-polynomial**	**SVM-RBF (γ=3.5)**	**KNN-k=3**	**SVM-KNN (RBF, γ=3.5 & k=3)**
Features count	8	8	8	8	8
No. clean training samples	10	10	10	10	10
No. outliers training samples	5	5	5	5	5
No. clean test samples	66	66	66	66	66
No. outliers test samples	5	5	5	5	5
No. true positives (TP)	2	4	4	4	4
No. true negatives (TN)	61	61	65	63	65
No. false positives (FP)	5	5	1	3	1
No. false negatives (FN)	3	1	1	1	1
Accuracy	0.8873	0.9155	0.9718	0.9437	0.9718
Detection rate	0.4	0.8	0.8	0.8	0.8
False Alarm rate	0.0758	0.0758	0.0152	0.0455	0.0152
True Negative rate	0.9242	0.9242	0.9848	0.9545	0.9848
Global Efficiency	0.6897	0.8780	0.9148	0.8965	0.9148

6 Discussion and Conclusion

This paper focuses on the outlier detection problem for multivariate data and proposes one measure (global efficiency) to classify different methods according their performance. Conventional outlier detection techniques are not suitable for multivariate data due to the special characteristics and resource limitations of the multivariate data. On the contrary the proposed algorithm is capable to detect outlier in multivariate data; therefore it offers a relatively more efficient way to detect multivariate data. The proposed approach takes into consideration various characteristics and features of the both methods (KNN and SVM) and evaluated with real life dataset (Milk data). The suggested global efficiency measure shows that our approach can achieve high accuracy rate for identifying outliers, and demonstrate the effectiveness of the proposed approach. In this paper, KSVM technique results are similar to SVM (RBF) results. But, giving the experimental results obtained by Shen and Lin [12], KSVM classifier has given a higher accuracy compared with the traditional SVM and conventional KNN. Based on their conclusions, the better classification performance of KSVM classifier compared with regular SVM classifier is because further information is carried after the training process is completed, as compared to SVM.

Besides, the number of attributes used in the training process will not have high effect to KSVM classifier than the two latter classifiers. In Mei et al., [14], KSVM can perform better than SVM only and KNN only, by having the prune function which can eliminate the mislabelled training samples effectively.

Therefore, we can conclude that KSVM technique is a good outlier detector and performs at least as good as other efficient methods. According to the discussion (given before) which shows that

KSVM is the best method for classification, consequently it can be also considered the best (Till now) in outlier detection which is a kind of classification.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Barnett V, Lewis T. Outliers in statistical data. John Wiley and Sons; 1994.

[2] Fawzy A, Mokhtar HMO, Hegazy O. Outliers detection and classification in wireless sensor networks. Egyptian Informatics Journal. 2013;157–164.

[3] Yang X, Latecki LJ, Pokrajac D. Outlier detection with globally optimal exemplar-based GMM. In: Proceedings of the 2009 SIAM International Conference on Data Mining (SDM), 145C154; 2009.

[4] Daudin JJ, Duby C, Trecourt P. Stability of principal component analysis studied by the bootstrap method. Statistics. 1988;19:241–258.

[5] Atkinson AC. Fast very robust methods for the detection of multiple outliers. Journal of the American Statistical Association. 1994;89:1329-1339.

[6] Caussinus H, Ruiz-Gazen A. Metrics for finding typical structures by means of principal components analysis, in data science and its applications, eds. Escoufier JC, Hayashi I. Japan: Academy Press; 1993.

[7] Rocke DM, Woodruff DL. Identification of outlier in multivariate data. Journal of the American Statistical Association. 1996;91(435):1047–1061.

[8] Mahalanobis PC. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India. 1936;2(1):49–55.

[9] Knorr EM, Ng RT. Algorithms for mining distance-based outliers in large datasets. In Proceedings of the 24rd International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc. 1998;392-403.

[10] Vapnik VN. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA; 1995.

[11] Rong L, Shiwei Y, Zhongzhi S. SVM-KNN classifier - a new method of improving the accuracy of SVM classifier. Acta Electronicasinica. 2002;30:5.

[12] Shen X, Lin Y. Gene expression data classification using SVM-KNN classifier, proceedings of the international symposium on intelligent multimedia, Video and Speech Processing, Oct. 20-22, Hong Kong, China. 2004;149-152.

[13] Zhang H, Berg AC, Maire M, Malik J. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition, proceedings of the IEEE computer society conference on computer vision and pattern recognition. June 17-22, Washington, DC, USA. 2006;2:2126-2136.

[14] Mei Z, Shen Q, Ye B. Hybridized KNN and SVM for gene expression data classification. Life Sci J. 2009;6:61-66.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=598&id=6&aid=5428