On Classification from the View of Outliers

Ching-an Hsiao, *Member, IAENG* and Hanxiong Chen

***Abstract*—Classification is the basis of cognition. Unlike other solutions, this study approaches it from the view of outliers. We present an expanding algorithm to detect outliers in univariate datasets, together with the underlying foundation. The expanding algorithm runs in a holistic way, making it a rather robust solution. Synthetic and real data experiments show its power. Furthermore, an application for multi-class problems leads to the introduction of the oscillator algorithm. The corresponding result implies the potential wide use of the expanding algorithm.**

***Index Terms*—Classification, outlier, expanding algorithm, sensitivity.**

I. Introduction

Classification is the basis of cognition. Of all the algorithms, neural networks, which simulate the function of neurons simply, have been proved to be a general and effective method. Currently, this method still appears to be robust and valuable [28, 29]. Unlike others, this study aims to find a basis for classification. We approach the problem from the view of outliers. One pattern is an outlier of another pattern, so outlier detection actually underlies the classification. The outlier problem can be traced to its origin in the middle of the eighteenth century, when the main discussion was about the justification to reject or retain an observation. "It is rather because…, that the loss in the accuracy of the experiment caused by throwing away a couple of good values is small compared to the loss caused by keeping even one bad value" [1]. There is still a great need for outlier detection in academia, industry, government, and research. From Peirce's old criterion [11] to current robust methods [9, 14], there are many different methods for detecting outliers. Some commonly used simple methods include Chauvenet's criterion [3], Boxplot [15], median and median absolute deviation (MAD) [5], and mean and standard variation. The problem is that the results from these seem to be inconsistent. It is as Pearson [10] stated: "even the best outlier-detection procedures can behave somewhat unpredictably, finding either more or fewer outliers in a data set than your eyes or other manual analyses might". This problem prompted our study. We approach the outlier detection problem in an ontological way, starting from the definition of an outlier. By analysing the nature of outliers, that is, inconsistency, we develop the concept of an integrated inconsistent rate (*IIR*) to express the outlier degree. Combined with Weber's Law, IIR, like humans, can distinguish outliers from normal values. Such classification is basic. This paper discusses related works and gives some examples to show the inconsistencies in commonly used methods in Section II. A new simple mechanism for detecting univariate outliers is presented in Section III. Section IV compares the new method with traditional ones using simulated and real data. We offer an extended discussion in Section V and our conclusions follow in Section VI.

C.A. Hsiao is with the Asian Water Environment Section, Asian Environment Research Group, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan (Tel/Fax: +81-29-850-2128, email: h_siao@hotmail.com).

H. Chen is with the Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan (e-mail: chx@cc.tsukuba.ac.jp).

II. Related Works

Due to manipulation errors by humans, system errors in sensors, or interference from unintended signals, some experimental data may differ greatly from the majority of the data and should be rejected. Traditional solutions approach the problem from the theory of probability. The old mean and standard deviation (σ) method assumes that data follow a normal distribution, and then uses a 95% (2σ) or 99.7% (3σ) boundary to identify "outliers". The Boxplot divides ordered data into four quartiles. Let the lower hinge (defined as the 25th percentile) be q1 and the upper hinge (the 75th percentile) be q3, then call the difference between them IQR (q3-q1), and any data outside the fence q1-1.5*IQR and q3+1.5*IQR are identified as outliers. The median and MAD method calculates the median and MADn of the data, where MADn = b*med$_i$|x_i-med$_j x_j$|, med$_i x_i$ is the median of data $\{x_1,\ldots,x_n\}$ and b=1.4826, and then uses median±kMADn to detect outliers. While the mean and standard deviation method uses a fixed coefficient (2 or 3) multiplied by the standard deviation (σ), Chauvenet's criterion uses a variable coefficient related to the number of data. In a recent work, Ross [13] suggested a return to the Peirce criterion, a forerunner of the probability approach. Rousseeuw presented various robust algorithms such as LMS and LTS [14], which were developed from the well-known least squares (LS) method. Differing from LS by using the idea of the least sum of squares as the regression estimator, LMS uses the least median of squares, while LTS uses the least trimmed squares. Thereafter, outliers are identified as those points that lie far away from the robust fit (a similar reasonable ratio, such as 1.5 for the Boxplot or k for the median and MAD method, is predetermined). Since there are so many algorithms, the problem arises of how to choose between them, especially in the face of contradiction. Turkey advised, "It is perfectly proper to use both classical and robust methods routinely, and only worry when they differ enough to matter. But when they differ, you should think hard" [9]. We

Table 1. Outliers detected by various methods

	Mean	Boxplot	MAD	Peirce's	LTS	LMS	IIR
ROSNER	none	40	40	40	40	40	40
BARNETT	none	none	949,951	none	949,951	949,951	949,951
GRUBBS1	none	596	584,596	596	578,584,596	578,584,596	596
GRUBBS3	none	none	none	none	2.02,2.22	2.02,2.22	2.02,2.22
CUSHNY	none	4.6	4.6	4.6	0,2.4,4.6	2.4,4.6	4.6

are still faced with the same difficulty, as can be seen by the examples here. In Table 1, for the median and MAD method (abbreviated as "MAD") we use k=3; the mean and standard deviation is abbreviated to "Mean"; and the robust LTS and LMS can be referred to the program "PROCESS" [14].

The following datasets were all taken from the web [16]. ROSNER contains 10 monthly diastolic blood pressure measurements; GRUBBS1 contains data on the strength of hard-drawn copper wire; GRUBBS3 consists of data on the percentage elongation of plastic material; and CUSHNY gives the difference in hours of sleep due to two different drugs used on ten patients. Obvious and common outliers are shown in bold italic font.

1. ROSNER: 90, 93, 86, 92, 95, 83, 75, ***40***, 88, 80
2. BARNETT: 3, 4, 7, 8, 10, ***949, 951***
3. GRUBBS1: 568, 570, 570, 570, 572, 572, 572, 578, 584, ***596***
4. GRUBBS3: 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94, 4.05, 4.11, 4.13
5. CUSHNY: 0, 0.8, 1, 1.2, 1.3. 1.3, 1.4, 1.8, 2.4, ***4.6***

Here, we note that using Peirce's criterion with BARNETT data, if we assume from the start that there are two doubtful observations, 949 and 951 are identified; while MAD used with GRUBBS3 data, detects 2.02 and 2.22 with k=2. Thus, different methods lead to different results. Is there no absolute outlier or no absolute detection method? The current situation is not satisfactory. Investigating the underlying foundation of one of the methods leads to even more confusion. When commenting on why 1.5 is used in the Boxplot method, Tukey said, "Because 1 is too small and 2 is too large" [23]. We should treat these solutions with the same attitude Hampel applied to the concept of an outlier, "without clear boundaries, nevertheless they are useful" [6]. The purpose of this study is to propose a simple, yet efficient and robust way of finding the "clear boundaries".

III. Ontological Criterion

A. Confirming the boundary

To confirm the boundary between normal data and outliers, a precise definition of an outlier is needed. We introduce the following well-known definition.

(An outlier is) an observation (or subset of observations), which appears to be inconsistent with the remainder of that set of data [2].

Barnett and Lewis [2] stated that "the phrase 'appears to be inconsistent' is crucial". Hawkins [7] also pointed out that an outlier is "an observation which deviates so much from other observations". Because inconsistency is the nature of an outlier and we cannot confirm such a characteristic from patterns outside of the data [8], we can only construct an inconsistent principle inside the data. Since inconsistency can be described as data from one position starting to appear very different from other ones (at least half of the whole), and distance is the best characteristic to express the difference of data, we developed the concept of an integrated inconsistent rate to detect outliers in univariate data.

Preliminaries

Let S denote an interval series $\{\delta_1, \delta_2, \ldots, \delta_N\}$ and $\Delta = \sum_{i=1}^{N} \delta_i$.

Three quantities are defined as follows:

Expansion ratio: $Er_i = N \times \delta_i / \Delta$

Inhibitory rate: $Ihr_i = \delta_i / (\delta_i - \max_{j<i}(\delta_j))$

Integrated inconsistent rate:

$$IIR_i = Er_i / Ihr_i = N \times (\delta_i - \max_{j<i}(\delta_j)) / \Delta$$

The expansion ratio expresses the ratio of the current interval (δ_i) to the average interval (Δ/N). The value of Er_i is 1 if there is no expansion in the current position compared with its "original" state. The greater the ratio is, the more likely it is to be the boundary separating outliers and normal data. The inhibitory rate is a modifying factor to the current *IIR* in terms of the former maximum interval and current interval. *Ihr* becomes infinity when δ_i is equal to $\max(\delta_j)$, in which case *IIR* is defined as 0. *IIR* takes both local and global characteristics into consideration, and thus gives an integrated inconsistent evaluation of the current interval with respect to others. A simple example to demonstrate the algorithm is a sequence of numbers with a common difference, such as {1, 2, 3, 4, 5, 6, 7}. For each number greater than 2, its expansion ratio, inhibitory rate, and integrated inconsistent rate are 1, $+\infty$, and 0, respectively.

The first element using *IIR* equal to or greater than c is confirmed as the boundary between outliers and normal values. Obviously, at least more than half the data should be normal, so outlier detection merely checks the remaining part. Suppose a dataset has outliers on the high value side, then the smallest value is the safest (normal) one. For such datasets, the following *Expanding Algorithm* (or *IIR algorithm*)

calculates the boundary distinguishing the outliers.

Algorithm 1:
Expanding Algorithm

Input: dataset D $\{d_1, d_2, \ldots, d_N\}$ with length N, and an adjustable threshold c
Output: outliers of dataset D
1. Sort D in ascending order; without loss of generality, we express the sorted D as
D=$\{d_{0:N}, d_{1:N}, \ldots, d_{N-1:N}\}$, where $d_{0:N} \le d_{1:N} \le \ldots \le d_{N-1:N}$
2. For i = 1 to N-1 do $\delta_i = d_{i:N} - d_{i-1:N}$
$\Delta = d_{N-1:N} - d_{0:N}$ //Δ is the sum of all δ_i
3. //Calculate Er, Ihr and IIR of $d_{i:N}$ ($i \ge 2$):
$\max_1 = 0$
For i = 2 to N-1 do
$\max_i = \max\{\max_{i-1}, \delta_{i-1}\}$
$Er_i = \delta_i \times (N-1)/\Delta$
$Ihr_i = \delta_i / (\delta_i - \max_i)$
$IIR_i = Er_i / Ihr_i$
4. $t = \min\{i \mid IIR_i > c \text{ and } i > N/2\}$
5. Output d_k when $d_k \ge d_{t:N}$

Obviously, each of the steps 2, 3, 4, and 5 in the above algorithm costs O(N), and thus the complexity of the algorithm is O(Nlog$N$), which comes from step 1.

In a similar way, the algorithm for a dataset with outliers on the lower value side can easily be designed. The algorithm, generalised as follows, solves the case where the safest point is in the middle.

Algorithm 1':
Expanding Algorithm

Input: dataset D $\{d_1, d_2, \ldots, d_N\}$ in ascending order
Output: outliers of dataset D

1. Set median set M=$\{d_{N/2}, d_{N/2+1}\}$ if N=even, or M=$\{d_{(N+1)/2}\}$ if N=odd.
2. Let the order of the minimum value of M be the left limit l, and the order of the maximum value be the right limit r. Initial value of l, r is l_0, r_0, where l_0 =N/2, r_0=N/2+1 if N is even, and $l_0 = r_0$=(N+1)/2 if N is odd.
3. Expanding median set M by step 4 till |M|=N/2+1 (N is even) or (N+1)/2 (N is odd).
4. If $(d_{r+1} - d_r) > (d_l - d_{l-1})$ then let left limit $l=l-1$
Otherwise, let right limit $r=r+1$
5. Calculate maxdelta=max$\{(d_i-d_{i-1}), (d_j-d_{j-1})\}$ ($i<l_0$, $j>r_0$ $d_i, d_j \in$ M)
6. Resume step 4 and calculate the following three parameters till IIR>=c or reaching all data (l=1 and r=N), c is the threshold.
To $i<l$:
$Er_i=(d_{i+1}-d_i)/(d_N-d_1)*(N-1)$
$Ihr_i=(d_{i+1}-d_i)/(d_{i+1}-d_i$-maxdelta)
$IIR_i=Er_i/Ihr_i$
If $IIR_i<c$
let $l=i$ and maxdelta=max{maxdelta, $(d_{i+1}-d_i)$}
to $j>r$:
$Er_j=(d_j-d_{j-1})/(d_N-d_1)*(N-1)$
$Ihr_j=(d_j-d_{j-1})/(d_j-d_{j-1}$-maxdelta)
$IIR_j=Er_j/Ihr_j$
if $IIR_j<c$
let $r=j$ and maxdelta=max{maxdelta, (d_j-d_{j-1})}
7. Accepted median set is normal set, that is, in [l,r], output others as outliers.

We demonstrate the execution of the algorithm in Table 2 using GRUBBS1 data. There is a total of ten data, so the middle two values, both 572, become members of the median set M in the beginning. Then the nearest neighbours of M are added in turn, until the size of M is 6. At this stage, maxdelta is calculated (step 5 in Algorithm 1'). Steps 6 to 9 in Table 2 correspond to step 6 in Algorithm 1'. When we reach 596, since its IIR is greater than c (1.81, refer to Section III.B), the boundary between normal data and outliers is confirmed.

B. Sensitivity

A sensitivity index IIR has been introduced in the Expanding Algorithm. It is a subjective parameter and can be deduced by Weber's law [18], which states that the ratio of the increment threshold (ΔI) to the background intensity (I) is a constant (K), i.e., $\frac{\Delta I}{I} = K$. All distinguishable quantities are related to this formulation. Suppose that three values deduced from the formulation: 0, I, $I+\Delta I$ are given. If we cannot tell I from $I+\Delta I$, the number 0 is a distinguishable quantity. Alternatively, we can make the transformation: $0, \Delta I, I+\Delta I$. If ΔI cannot be sensed, $I+\Delta I$ differs (sensible) from the others (0 and ΔI). We express the three values as two intervals (N=2), i.e., $\delta_1 = \Delta I$, and $\delta_2 = I$.

Table 2. Outlier detection process for GRUBBS1 data

Step	Members of median set added	Er	Ihr	IIR	maxdelta
1	572, 572				
2	572				
3	570				
4	570				
5	570				2
6	568	0.64	∞	0	2
7	578	1.93	1.5	1.29	6
8	584	1.93	∞	0	6
9	596 (outlier)	3.86	2	1.93	

Table 3. Typical *IIR* in the three values system

K	IIR
0	2
0.01	1.96
0.05	1.81
0.1	1.64
1	0

According to the preliminaries in Section III.A, we obtain three parameters corresponding to interval *I* (the case when *i=2*).

$$Er = \frac{2 \times I}{I + \Delta I} = \frac{2}{1 + K}$$

$$Ihr = \frac{I}{I - \Delta I} = \frac{1}{1 - K}$$

$$IIR = \frac{Er}{Ihr} = \frac{2 \times (1 - K)}{1 + K}$$

We have a reasonable *K* in (0, 1), with the corresponding typical *IIR* given in Table 3. The threshold *c* in the Expanding Algorithm is assigned to 1.81 in this paper.

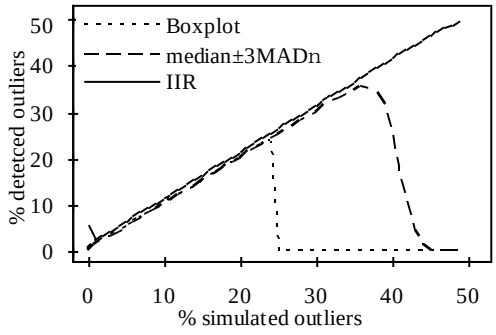
IV. Experiments

For performing outlier tests, one approach is to use an outlier-generating model that allows a small number of observations from a random sample to come from a distribution G differing from the target distribution F [2]. Observations not from F are called contaminants. The object of finding outliers is to detect the contaminants.

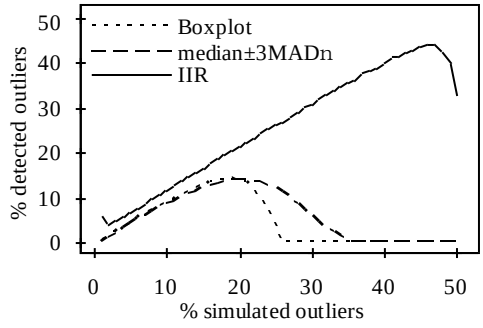
Reimann et al. [12] compared the three-sigma rule, the Boxplot, and the MAD method. Here, we compare the same methods, but replace the three-sigma rule with our IIR algorithm. In the first simulation, both F and G were normal distributions, with means of 0 and 10, respectively, and standard deviation of 1. A fixed sample size of N=500 was used, of which different percentages (0-49% step 1%) were outliers drawn from G. In the comparison of the Boxplot, the MAD method, and the IIR algorithm, each simulation was replicated 1000 times and the average percentage of detected outliers computed. The results are shown in Fig. 1(a). In the second simulation, the mean of the G distribution was changed to 5, with the results depicted in Fig. 1(b).

According to Fig. 1(a), the Boxplot and MAD method perform well. Nevertheless, at 25% outliers, the Boxplot breaks down; and at 37% outliers, the MAD method breaks down. The IIR algorithm performs consistently, always overestimating the number of outliers slightly. According to Fig. 1(b), the Boxplot breaks down at 19% outliers and the MAD breaks down at 20%, whereas our IIR algorithm retains its efficient execution until 47%. In fact, the distance between distributions of means 0 and 5 is so close that the separation of normal values and outliers is no longer clear. Considering that the IIR algorithm can still detect 32.5% of outliers with a contamination percentage of up to 49%, clearly shows that it is indeed robust.

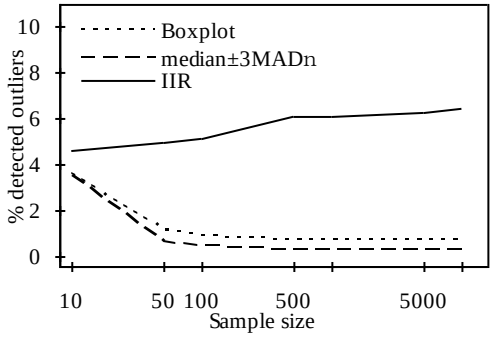
A further experiment explains why the IIR algorithm



(a) Distributions with means 0 and 10



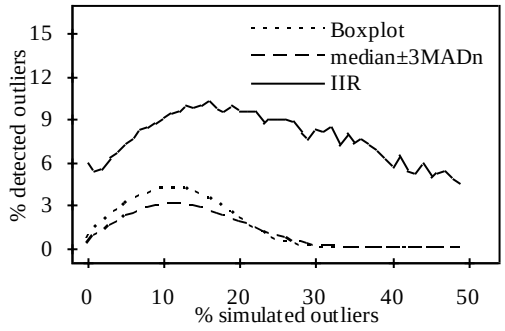
(b) Distributions with means 0 and 5



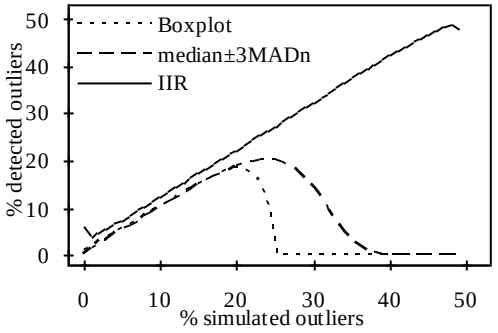
(c) Standard normal distribution

Fig. 1. Average percentage of outliers detected by three methods

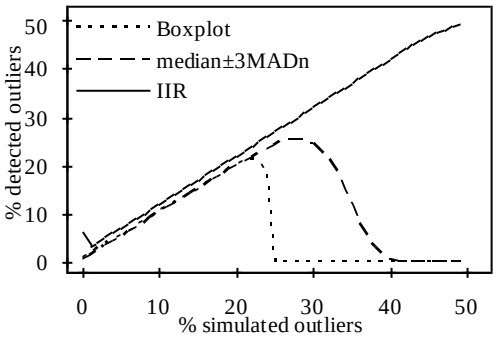
overestimates outliers. For simulated standard normal distributions with sample sizes 10, 50, 100, 500, 1000, 5000 and 10000, we computed the percentage of detected outliers for the three methods. Each sample size was replicated 1000 times and the average results are shown in Fig. 1(c). As the sample size increases, the Boxplot and the MAD method tend to detect outliers less than 1% of the time, while the IIR algorithm appears robust and its detection ability increases slightly. In theory, the probability of the appearance of a deviation point increases with an increase in sample size. But



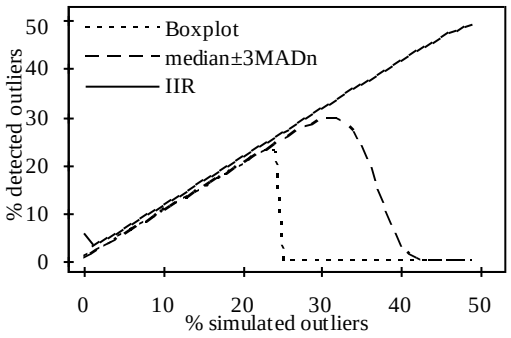
(a) Distributions with means 0 and 3



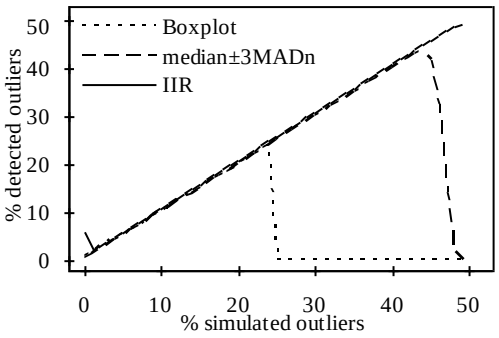
(b) Distributions with means 0 and 6



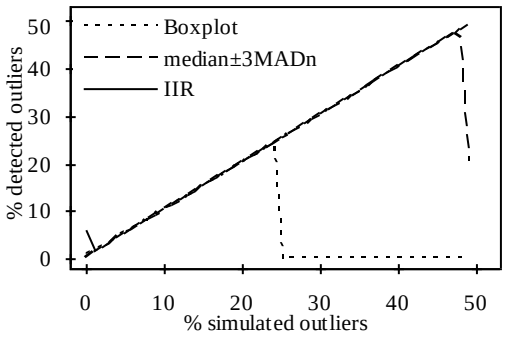
(c) Distributions with means 0 and 7



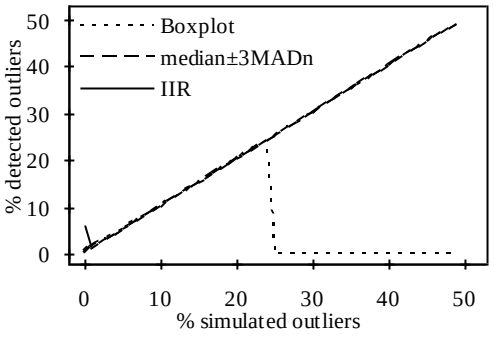
(d) Distributions with means 0 and 8



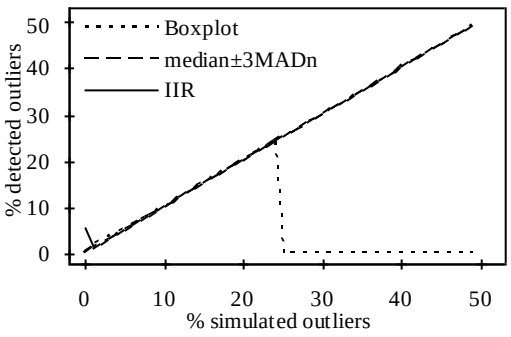
(e) Distributions with means 0 and 15



(f) Distributions with means 0 and 20



(g) Distributions with means 0 and 25

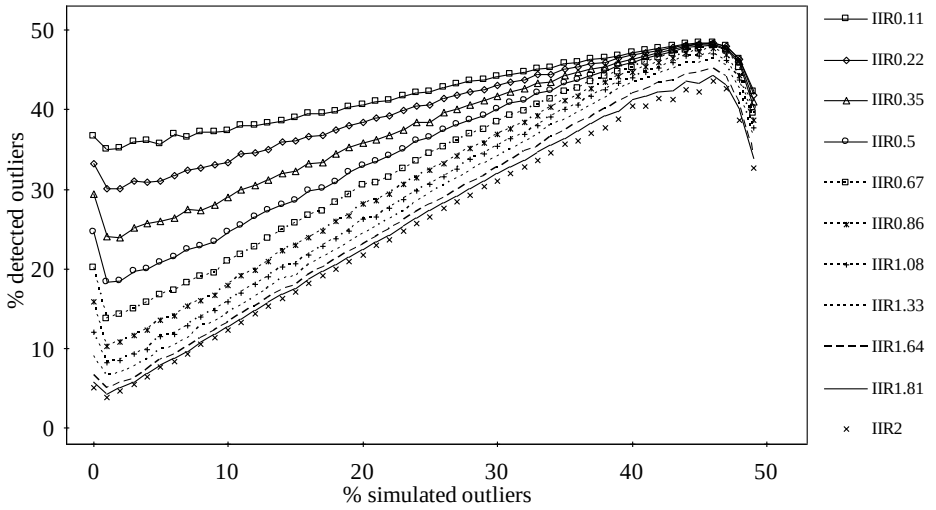


(h) Distributions with means 0 and 30

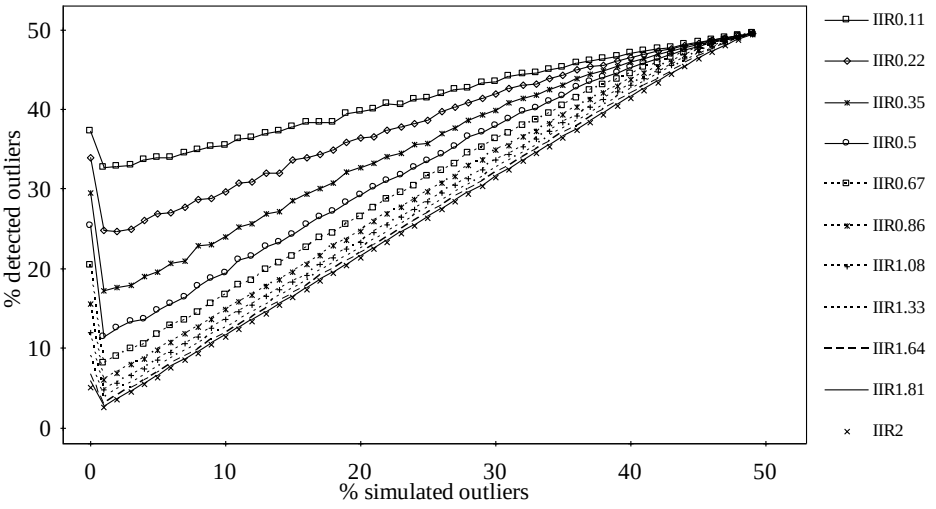
Fig. 2. Average detection percentage for various distributions

appearance does not mean consistency; and the IIR algorithm can detect such inconsistency, while the other two methods seem to fail. The reason that the IIR algorithm always overestimates outliers with smaller sample sizes is that the IIR algorithm not only detects contaminants, but also detects outliers in the target distribution itself.

To illustrate the character of the IIR algorithm better, we extend the previous experiment to make additional



(a) Distributions with means 0 and 5



(b) Distributions with means 0 and 10

Fig. 3. Detected percentage by different IIR

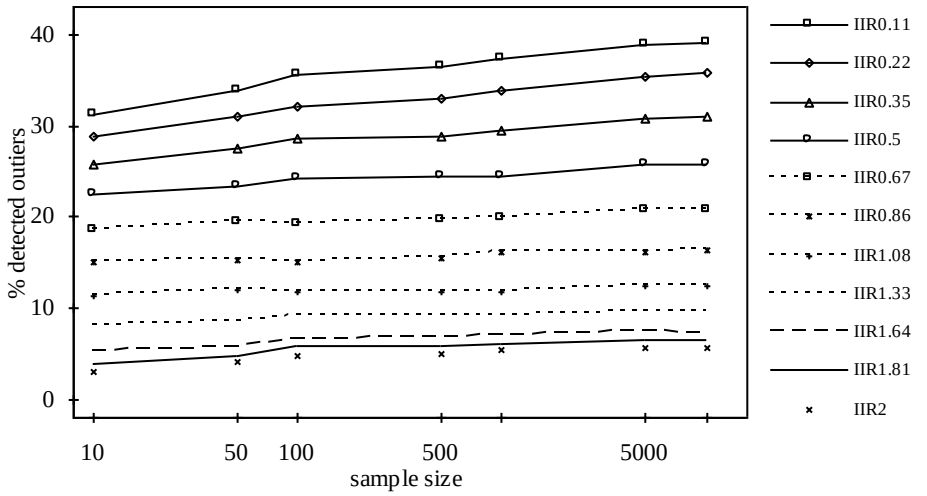


Fig. 4. Different IIR cases for standard normal distribution

comparisons. Contaminants are extended to distributions with means 3, 6, 7, 8, 15, 20, 25, and 30 with the same standard deviation 1. Fig. 2 shows the results. The Boxplot appears consistent within its capability. With a mean greater than or equal to 25, the MAD method no longer breaks down, while the Expanding Algorithm achieves this at mean 7. At mean 30, the average interval between the Expanding Algorithm and the MAD method (in this case, nearly all contaminants are detected) decreases to 0.4% (simulated outliers are from 1% to 49%), a tiny value extracted from the target itself. As a comparison, the average interval between the Expanding Algorithms at mean 7 and mean 30 is 1.64%, which is fairly robust.

In addition, we compared the results using different sensitivity thresholds. We used 0.11, 0.22, 0.35, 0.5, 0.67, 0.86, 1.08, 1.33, 1.64, 1.81, and 2 based on distributions with means 0 and 5, and distributions with means 0 and 10, all with standard deviation 1. The results are shown in Fig. 3. We observe that different sensitivity degrees produce different results, but the trend and breakdown points are all similar. Instances with average error less than 5% (from 0% to 49%) are 1.33, 1.64, 1.81, and 2. This again confirms the consistency of the Expanding Algorithm even with different thresholds, and a threshold around 1.81 concurs with our initial sense. Thresholds less than 1 are considered too sensitive. Another experiment, with results depicted in Fig. 4, has the same conclusion. For different sample sizes of a standard normal distribution, different IIR thresholds have different sensitivity, with between 1.33 and 2 being acceptable.

This paper also makes use of real data [17], in which "there is little room for argument about what the outliers are" [4]. The data consist of 2001 measurements of radiation taken from a balloon about 30-40 kilometres above the earth's surface. As reported by the Hampel inward procedure, 396 observations are identified as outliers (normal observations are all between y=±0.1). All the obvious outliers are identified, leaving only a few doubtful cases of no great importance. In this case, median±3MADn detected 347 outliers, and Boxplot 297 outliers. LST detected 440 outliers (normal values in [-0.065, 0.089]), while LMS detected 428 outliers (normal values in [-0.068, 0.098]). The IIR algorithm detected 398 outliers (normal values in [-0.084, 0.092]). Compared with the results of the Hampel inward procedure (normal values are in [-0.083, 0.1]), and considering outside neighbours of -0.083 (three -0.084s and two -0.091s), and outside neighbours of 0.092 (0.097, 0.098, 0.099, 0.099, 0.1, and 0.111), the IIR algorithm is found to have better location ability. It is obvious that the other methods were not able to take into account the local properties, making it difficult for them to capture correctly the exact boundaries (locally related).

Table 4. Observations of the vertical semi-diameter of Venus

-0.30	+0.48	+0.63	-0.22	+0.18
-0.44	-0.24	-0.13	-0.05	+0.39
+1.01	+0.06	-1.40	+0.20	+0.10

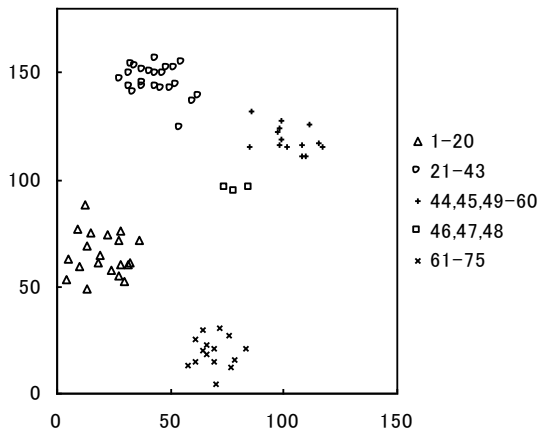


Fig. 5. Ruspini data (five clusters by Oscillator Algorithm)

The results of applying the IIR algorithm to the datasets in Section II are given in Table 1. The same positive conclusion can be drawn here.

V. Discussion

In this section, we discuss a rather famous set of observations and then give an example of a multi-class case.

The classic set (Table 4.) consists of a sample of 15 observations of the vertical semi-diameter of Venus made by Lieutenant Herndon using the meridian-circle at Washington in 1846 [11].

Peirce applied his criterion and rejected two observations, +1.01 and -1.40 [11]. Later, Gould recalculated Peirce's criterion with increased precision and identified only +1.01 [24]. The Boxplot and the MAD method mentioned above all label -1.40 as the only outlier. LMS and LTS both detect two outliers, +1.01 and -1.40. Grubbs confirmed -1.40 to be rejected and +1.01 to be retained at the 5% level [25]. Tietjen and Moore used a one variable Grubbs-type statistic to reject both -1.40 and +1.01, and declared their method to have a real significance level of 0.05 [26]. Barnett et al. [2] found even -1.40 not to be an outlier, although they used mismatched data. Nevertheless, the problem lies with +1.01, and not with -1.40.

If we use Tietjen and Moore's method with CUSHNY data (Section II), we obtain E_2=0.128, E_3=0.044, E_4=0.026. These values are all smaller than the corresponding 5% critical values of 0.172, 0.083 and 0.037. Thus, 4.6, 0, 2.4, and 0.8 should all be labelled as outliers. Their evaluation of the case is exactly so: using the appropriate value of k for E_k is important, otherwise an error occurs. However, can a decision be made before it is processed? Before we suggest an answer, we analyse this example using the IIR algorithm.

Using 1.81 as the sensitivity threshold, we only identify -1.40 as an outlier. What about +1.01? Its *IIR* is 1.10, which means it is detected at K=0.29. The next larger value of IIR is at 0.39, the *IIR* of which is 0.29 and the corresponding K is 0.75, which is distant from 0.29 and in quite a different "sense" level, where a smaller K means a more sensitive system. Using a different sensitivity, we possess different

knowledge. The IIR algorithm is a consistent method. Further description of outliers can be found in the works by Hsiao et al. [8, 27].

Although the above algorithm solves the problem of two classes, we show, by means of an example, that the Expanding Algorithm can be applied to problems with multiple classes. The Ruspini dataset, consisting of 75 points (Fig. 5) in four groups, is commonly used to illustrate clustering techniques [22]. Clustering is one of the classic problems in machine learning. A popular method is k-means clustering [19, 20]. Although its simplicity and speed are very appealing in practice, it offers no guarantees of accuracy. Furthermore, solving the problem exactly is NP-hard [21]. Like k-means, most algorithms use the center point to represent a cluster, and each element is classified according to its distance from the closest center. In reality, this is not always so, and the absolute center is not necessary (this does not imply that the center point is useless). Based on this observation, a new method is presented to cluster Ruspini data.

Given a Ruspini dataset D $\{d_1, d_2, \ldots, d_{75}\}$, with each point as a cell.

Oscillator Algorithm:

1. Calculate distances between any two points d_i and d_j.
2. For any point d_i, arrange its distance series (to other points) in ascending order.
3. Calculate the series of any i using the Expanding Algorithm under the condition that the safest point is the first one and at least three points are included for more than two classes to exist. From this we obtain 75 clustering sets.
4. Randomly choose one point as a seed with firing intensity 1, and let the intensities of other points be 0.
5. Any partner (clustering member) of the firing cell can receive its stimulus and thus begin to fire with the same intensity, while others receive an identical negative input.
6. Repeat step 5 until all cells remain unchanged or are fully charged (including negatively charged).
7. Cluster all cells with positive firing in one cluster.
8. Repeat steps 4 to 7 on the remaining points.
9. Alternative approach: combine all the results of each cell, and determine clusters.

Fig. 5 shows the clustering outcome using the Oscillator Algorithm. The data are clustered into five groups. Considering the small scale of cluster 4 (46-48), we can easily merge it with its nearest neighbour - cluster 3. In this case, the result remains the same as designed. However, without any extra information or restriction, cluster 4 could also be treated as outliers.

Table 5 gives the detailed results of the Oscillator Algorithm. Each cell was chosen as a seed in turn, and two different results were achieved. One matches the result of five clusters; in the other case, cells remain silent, which means that the corresponding cells have no way of obtaining resonance. From the results, we can see the effect of the Oscillator Algorithm, based completely on uncertainty and being just what we need to model the mind.

Table 5. Clustering summary from the view of each element

Clusters and Included Elements	Number of elements with correct clustering	Silent elements	Probability for correct clustering
1 (1-20)	18	17, 20	90%
2 (21-43)	21	41, 42	91%
3 (44,45, 49-60)	11	44,45,58	79%
4 (46,47,48)	2	46	67%
5 (61-75)	15	-	100%

VI. Conclusion

This paper is concerned with the outlier detection problem for univariate data, which can also be viewed as a primary pattern classification problem. The Expanding Algorithm is presented, together with three clearly defined parameters (Er, Ihr, IIR) to express the degree of the outlier, which clarifies the related problem in a certain way. Furthermore, a sensitivity index based on Weber's law is combined seamlessly to create an effective system. Experiments using both simulated and real data show the robustness of the system. A deeper relation between patterns and outliers can be found in [8] [27], where a general framework was constructed to describe and calculate patterns – a key factor for intelligence. In this paper, an extended application is also discussed for multi-class problems, and the result strengthens the conclusion in [27]. Above all, any classification can be treated as a type of distinction between numbers that correspond to the characteristics or features of things. Distinction is the foundation of human cognition. Indistinguishable items are classified into one group, and the difference within a class is less than that between classes. The underlying distinction or inconsistency can be expressed simply and well using IIR, which takes into account both the whole and the detail. The ability to distinguish correlates with the level of IIR, i.e., the different thresholds of IIR lead to different precision results. This condition mimics human thought, and thus the Expanding Algorithm based on the inconsistency principle can be widely used in classification. It is also expected that this method could result in an effective mind model when combined with previous and future works.

Acknowledgment

The authors are grateful to Paul F. Velleman and Laurie Divies. They would also like to thank the reviewers for their valuable comments.

References

[1] V. Barnett and T. Lewis, Outliers in statistical data, John Wiley & Sons, 1978, pp1
[2] V. Barnett and T. Lewis, Outliers in statistical data, Wiley & Sons, 3rd edition, 1994

[3] W. Chauvenet, A Manual of Spherical and Practical Astronomy V.II, Lippincott, Philadelphia, 1st Ed, 1863
[4] L. Davies and U. Gather, "Rejoinder", Journal of the American statistical association, Vol.88, No.423, 1993
[5] F. Hampel, "The influence curve and its role in robust estimation", Journal of the American Statistical Association, 69, 383-393, 1974
[6] F. Hampel, "Robust statistics: a brief introduction and overview", Invited talk in the Symposium "Robust Statistics and Fuzzy Techniques in Geodesy and GIS", Zurich, 2001
[7] D.M. Hawkins, Identification of Outliers, Chapman and Hall, 1980
[8] C.A. Hsiao, H. Chen, K. Furuse and N. Ohbo, "A relative deviation detection for time series data based on Equality", Proceedings of International Multi-Conference on Engineer and Computer Scientist Vol. 1, pp. 511-516, 2008
[9] R.A. Maronna, R.D. Martin and V.J. Yohai, Robust Statistics: Theory and Methods, John Wiley & Sons, 2006
[10] R. Pearson, "Scrub data with scale-invariant nonlinear digital filters", EDN 71, January 24, 2002
[11] B. Peirce, "Criterion for the rejection of doubtful observations", Astronomical Journal II, 45, 161-163, 1852
[12] C. Reimann, P. Filzmoser and R.G. Garrett, "Background and threshold: critical comparison of methods of determination", Science of the Total Environment, 346, 1-16, 2005
[13] S.M. Ross, "Peirce's criterion for the elimination of suspect experimental data", Journal of Engineering technology, Fall, 2003
[14] P.J. Rousseeuw and A.M. Leroy, Robust regression and outlier detection, John Wiley & Son, 1987
[15] J.W. Tukey, Exploratory data analysis, Addison-Wesley, 1977
[16] Datasets package "PROGESS", available: http://www.agoras.ua.ac.be, retrieved July 21, 2009
[17] Balloon residuals data, available: http://lib.stat.cmu.edu/datasets/balloon, retrieved July 21, 2009
[18] E. H. Weber, *The Sense of Touch*, English translation by Ross, H. E. & Murray, D. J. Academic Press, London. New York. San Francisco, 1978
[19] S. P. Lloyd, Least squares quantization in pcm. *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129-136, 1982
[20] D. Arthur and S. Vassilvitskii, k-means++ The Advantages of Careful Seeding, Symposium on Discrete Algorithms (SODA). 2007
[21] Drineas, P., Frieze, A., Kannan, R., Vempala, S. and Vinay, V. Clustering large graphs via the singular value decomposition. *Mach. Learn*., vol. 56, pp. 9-33. 2004.
[22] E. H. Ruspini, Numerical methods for fuzzy clustering. *Inform Sci*, vol. 2, pp. 319-350, 1970
[23] Richard D. De Veaus, Paul F. Velleman, Intro Stats, Addison Wesley, 2003
[24] B.A. Gould, "On Peirces's criterion for the rejection of doubtful observations, with tables for facilitating its application", Astronomical Journal IV, 83, 81-87, 1855
[25] F. E. Grubbs, "Procedures for detecting outlying observations in samples", Technometrics, vol. 11, No.1, 1969
[26] G. L. Tietjen and R. H. Moore, "Some Grubbs-type statistics for the detection of several outliers". Technometrics, 14, 583-597, 1972
[27] C.A. Hsiao, H. Chen, K. Furuse, and N. Ohbo, Figure and ground: a complete approach to outlier detection. *IAENG Transactions on Engineering Technologies Vol. 1*, Ao, S.-L., Chan, A. H.-S., Katagiri, H., Castillo, O. & Xu, L. (Eds.), 70-81, American Institute of Physics, New York, 2009
[28] R. Sukanesh and R. Harikuma, "A Patient Specific Neural Networks (MLP) for Optimization of Fuzzy Outputs in Classification of Epilepsy Risk Levels from EEG Signals", Engineering Letters, 13(2), 50-56, 2006.
[29] Pérez Aguila, R, "Automatic Segmentation and Classification of Computed Tomography Brain Images: An Approach Using One-Dimensional Kohonen Networks", IAENG International Journal of Computer Science, 37(1), 27-35, 2010