A Unifying Method for Outlier and Change Detection from Data Streams Based on Local Polynomial Fitting

Zhi Li[1], Hong Ma[1], and Yongbing Mei[2]

[1] Department of Mathematics, Sichuan University, Chengdu, 610064, China
zhili_mail@163.com
[2] Southwest China Institute of Electronic Technology, Chengdu, 610036, China
meiybmail@sina.com

Abstract. Online detection of outliers and change points from a data stream are two very exciting topics in the area of data mining. This paper explores the relationship between these two issues, and presents a unifying method for dealing with both of them. Previous approaches often use parametric techniques and try to give exact results. In contrast, we present a nonparametric method based on local polynomial fitting, and give approximate results by fuzzy partition and decision. In order to measure the possibility of being an outlier and a change point, two novel score functions are defined based on the forward and backward prediction errors. The proposed method can detect outliers and changes simultaneously, and can distinguish between them. Comparing to the conventional parametric approaches, our method is more convenient for implementation, and more appropriate for online and interactive data mining. Simulation results confirm the effectiveness of the proposed method.

Keywords: Data stream, outlier, change point, data mining, local polynomial fitting, fuzzy partition.

1 Introduction

As there is a growing number of emerging applications of data streams, mining of data streams is becoming increasingly important. Recent research indicates that online mining of the changes in data streams is one of the core issues with applications in event detection and activity monitoring [1]. On the other hand, outlier detection is also a popular issue in data mining, which is closely related to fraud detection, abnormality discovery and intrusion detection [2]. In the previous literature, outlier detection and change detection are often derived from respective problems and are addressed independently. Both statistical methods and fuzzy approaches have been employed to solve these two issues, such as methods based on regression analysis, hidden Markov model (HMM), hierarchical Bayesian model and fuzzy clustering, fuzzy entropy principle [7], etc.

However, the outliers and change points often exist simultaneously in real data streams. It's necessary to design a unifying method to detect the outliers and change points simultaneously. Thus, in this paper, we explore the relationship between outlier and change detection, and deal with them together. Intuitively, an outlier is a point

Z.-H. Zhou, H. Li, and Q. Yang (Eds.): PAKDD 2007, LNAI 4426, pp. 150–161, 2007.

largely deviating from the holistic regularity, while a change point is a point from which the holistic regularity changes. Although outlier and change are two different concepts, they can be unified based on a probabilistic model. When a data stream is modeled as a time series with some probabilistic structure, both outliers and changes can be defined by the variation of statistical regularity, and the only difference is the variation kind. In [4], a unifying framework for mining outliers and changes was proposed, but the two issues were dealt with at two different stages. In [5], we have developed this work into a one-stage framework based on the forward and backward predictions. However, these two methods need pre-selected parametric models, and the parameters must be estimated adaptively in real time implementation. These will increase the difficulty in application, and are the drawbacks of all the parametric methods.

In this paper, we propose a nonparametric unifying method for online mining outliers and changes from data streams. The data stream herein is modeled as a time series with some probabilistic structure. An outlier is defined as a point with both small forward and backward conditional density, while a change is a point with small forward conditional density and large backward conditional density. In order to measure the possibility of being an outlier and a change point, we define two score functions based on the forward and backward prediction errors. Unlike parametric approaches, all predictions are estimated using the local polynomial fitting technique [6] which does not need parameter estimation, but approximates the predictions by fitting a polynomial using the local data around the testing point. This nonparametric method provides many advantages. For example, there's no need to determine the type of the time series model. The prediction accuracy will not be affected by the parameter estimation error. It is appropriate to both the linear and nonlinear time series, which is difficult for parametric methods.

Approaches proposed in the previous literature often try to give an exact partition among outliers, changes, and normal points. However, exact answers from data streams are often too expensive, and approximate answers are acceptable [3]. So in this paper, fuzzy partition and decision approaches are used to alarm possible outliers and changes. The magnitude of the possibility is visualized by the values of membership functions based on which people can make their own decisions. Thus, we believe that our method will be more effective in online and interactive mining of outliers and changes.

The rest of the paper is organized as follows: In Section 2, we formulate the problem of outlier and change detection, and give formal definition of outlier and change point. We give a brief introduction to the local polynomial fitting technique in Section 3, and present the unifying nonparametric outlier and change detection method in Section 4. Simulation results on several data sets are provided in Section 5 and a section of conclusion follows.

2 Problem Formulation

In this section, we will formulate the problem of outlier and change detection from the statistical point of view. While the term "outliers" or "changes" sounds general and intuitive, it is far from easy to define them. One natural description is that an outlier is a point largely deviates from the holistic regularity, while a change point is a point from which the holistic regularity changes. Although this description is inexplicit, it suggests something common between outlier and change. The holistic regularity varies at both outlier and change point, and only the type of the variation is different. Therefore, detection of outliers and changes is to find the variations of the regularity and distinguish between the different types. Considering a data stream $\{x_t : t = 1, 2, \cdots\}$, if it is modeled with some probabilistic structure, its conditional probability distribution can be incrementally learned from the data stream every time a datum x_t is input. That means we can learn the statistical regularity of the data stream adaptively and find the variations.

We model the real data stream $\{x_t\}$ as a local stationary time-series. Here, each x_t is regarded as a continuous random variable. We use the notation $p(x_t \mid x_{t-L}^{t-1})$ to represent the conditional density of x_t given by $x_{t-L}, \cdots, x_{t-2}, x_{t-1}$, and call it *forward conditional density*. Similarly, the notation $p(x_t \mid x_{t+1}^{t+L})$ is used to represent the conditional density of x_t given by $x_{t+1}, x_{t+2}, \cdots, x_{t+L}$, and is named *backward conditional density*. Then, the formal definition of outlier and change point is given as follows: an outlier x_t is a point with small $p(x_t \mid x_{t-L}^{t-1})$ and small $p(x_t \mid x_{t+1}^{t+L})$, while a change point x_t is a point with small $p(x_t \mid x_{t-L}^{t-1})$ and large $p(x_t \mid x_{t+1}^{t+L})$. Here, we are only interested in the sudden changes.

Now, some criterions should be selected to measure the possibility of being an outlier and a change point. In many previous literature, parametric time-series model is employed, and the form of the conditional density function is pre-decided. Then, the unknown parameters in the conditional density function can be estimated adaptively from $\{x_t\}$. Two score functions are often used as criterions in the parametric approaches [4], [5]. One is based on logarithmic loss:

$$Score(x_t) = -\log p(x_t \mid x_{t-L}^{t-1}, \theta_{t-1}), \tag{1}$$

where $p(x_t \mid x_{t-L}^{t-1}, \theta_{t-1})$ is the estimated parametric conditional density function at time point $t-1$. Another one is based on quadratic loss:

$$Score(x_t) = (x_t - \hat{x}_t)^2, \tag{2}$$

where, $\hat{x}_t$ denotes the prediction for x_t given $x_{t-L}, \cdots, x_{t-2}, x_{t-1}$ based on the estimated conditional density function as follows:

$$\hat{x}_t = E[x_t \mid x_{t-L}^{t-1}] = \int x p(x \mid x_{t-L}^{t-1}, \theta_{t-1}) dx. \tag{3}$$

However, the estimation for the parametric conditional density function is based on parametric modeling and enough data. In online data mining, only limited data are available for parametric modeling. Thus the modeling biases may arise with high probability and the detection accuracy will be degraded. Moreover, many data in applications exhibit nonlinear features that require nonlinear models to describe. However, beyond the linear parametric models, there are infinitely many nonlinear forms that can be explored. This would be a daunting task for any analysts to try one model after another. In addition, both of the two score functions only consider the forward conditional density, which can not distinguish between the outliers and changes. Thus, in this paper, we propose two novel score functions based on the forward and backward prediction errors (see Section 4.1), and employ a simpler and effective nonparametric approach, the local polynomial fitting, to calculate the predictions.

3 Local Polynomial Fitting

Local polynomial fitting is a widely used nonparametric technique. It possesses various nice statistical properties [6]. Consider a bivariate sequence $\{(X_t, Y_t): t = 1, \cdots, N\}$ that can be regarded as a realization from a stationary time series. We are interested in estimating Y_t by X_t, and the best estimation of Y_t based on $X_t = x$ is the conditional expectation of Y_t given $X_t = x$. Define a regression function in the following form:

$$m(x) = E(Y_t \mid X_t = x), \tag{4}$$

then Y_t can be expressed as follows:

$$Y_t = m(x) + \sigma(x)\varepsilon_t, \tag{5}$$

where $\sigma^2(x) = Var(Y_t \mid X_t = x)$, and ε_t is a random variable that satisfies $E(\varepsilon_t \mid X_t) = 0$, $Var(\varepsilon_t \mid X_t) = 1$.

Denote an arbitrary value of the regression function by $m(x_0)$. Local polynomial fitting is a method for estimating $m(x_0)$. Since the form of $m(x)$ is not specified, a remote data point from x_0 provides very little information about $m(x_0)$. Hence, we can only use the local data around x_0. Assume that $m(x)$ has the $(p+1)$ derivative at the point x_0. By Taylor's expansion, for x in the local neighborhood of x_0, we have the *local model*:

$$m(x) \approx \sum_{j=0}^{p} \beta_j (x - x_0)^j, \tag{6}$$

where $\beta_j = m^{(j)}(x_0)/j!$ and are called *local parameters*. One can estimate the local parameters by minimizing

$$\sum_{t=1}^{N} \left\{ Y_t - \sum_{j=0}^{p} \beta_j (X_t - x_0)^j \right\}^2 K_h(X_t - x_0). \tag{7}$$

The weight function $K_h(\cdot)$ is defined as $K_h(\cdot) \triangleq K(\cdot/h)/h$, where $K(\cdot)$ is a *kernel function* and h is a window bandwidth controlling the size of the local area. The parameter p is named *fitting order*. Formula (7) means the local parameters are estimated by fitting the local model (6) using the local data in the area $[x_0 - h,\ x_0 + h]$. Since $\hat{m}(x_0) = \hat{\beta}_0$, estimate for $m(x_0)$ is actually the weight least square (WLS) solution to the minimizing problem of (7).

4 Outlier and Change Detection

As mentioned before, both of the two score functions (1) and (2) need parametric estimation of the conditional density function, and can not distinguish between the outliers and changes. So in this section, we define two novel score functions based on the forward and backward prediction errors. These two scores are then used to alarm possible outliers and changes based on fuzzy partition and decision. If the prediction of x_t is regarded as a regression function, it can be calculated by local polynomial fitting without parametric modeling. Thus, the outliers and changes can be detected by nonparametric techniques.

4.1 Forward and Backward Scores

We first consider two bivariate sequences $\{(x_{t-L}, x_{t-L+1}), \cdots, (x_{t-2}, x_{t-1})\}$ and $\{(x_{t+2}, x_{t+1}), \cdots, (x_{t+L}, x_{t+L-1})\}$. Then two regression functions can be defined as the *forward prediction* of x_t which means prediction of x_t given by x_{t-1}:

$$m_f(x) = E(x_t \mid x_{t-1} = x), \tag{8}$$

and the *backward prediction* of x_t which means prediction of x_t given by x_{t+1}:

$$m_b(x) = E(x_t \mid x_{t+1} = x). \tag{9}$$

Similar local models can be defined as (6), and *forward and backward local parameters* can be defined as

$$\beta_j^{(f)} = m_f^{(j)}(x_{t-1})/j!\text{, and } \beta_j^{(b)} = m_b^{(j)}(x_{t+1})/j!,\ \ j = 0, \cdots, p\,. \tag{10}$$

Fitting the local models using the forward data $\{x_{t-L}, \cdots, x_{t-2}\}$ and the backward data $\{x_{t+2}, \cdots, x_{t+L}\}$ respectively, estimates for the forward and backward predictions can be obtained:

$$\hat{\beta}^{(f)} = (X_f^T W_f X_f)^{-1} X_f^T W_f y_f, \tag{11}$$

and

$$\hat{\beta}^{(b)} = (X_b^T W_b X_b)^{-1} X_b^T W_b y_b. \tag{12}$$

where

$$y_f = (x_{t-L+1}, \ldots, x_{t-1})^T,\ y_b = (x_{t+1}, \ldots, x_{t+L-1})^T,$$
$$W_f = diag(K_h(x_{t-L} - x_{t-1}), \cdots, K_h(x_{t-2} - x_{t-1})),$$
$$W_b = diag(K_h(x_{t+2} - x_{t+1}), \cdots, K_h(x_{t+L} - x_{t+1})),$$
$$X_f = \begin{pmatrix} 1 & (x_{t-L} - x_{t-1}) & \cdots & (x_{t-L} - x_{t-1})^p \\ \vdots & \vdots & & \vdots \\ 1 & (x_{t-2} - x_{t-1}) & \cdots & (x_{t-2} - x_{t-1})^p \end{pmatrix},$$
$$X_b = \begin{pmatrix} 1 & (x_{t+2} - x_{t+1}) & \cdots & (x_{t+2} - x_{t+1})^p \\ \vdots & \vdots & & \vdots \\ 1 & (x_{t+L} - x_{t+1}) & \cdots & (x_{t+L} - x_{t+1})^p \end{pmatrix}.$$

Thus the forward and backward predictions of x_t are $\hat{\beta}_0^{(f)}$ and $\hat{\beta}_0^{(b)}$, based on which two novel score functions are defined to measure the possibility of being an outlier and a change point. One is *Forward Score*:

$$Score_f(x_t) = (x_t - \hat{\beta}_0^{(f)})^2 \big/ \hat{\sigma}_f^2, \tag{13}$$

another one is *Backward Score*:

$$Score_b(x_t) = (x_t - \hat{\beta}_0^{(b)})^2 \big/ \hat{\sigma}_b^2. \tag{14}$$

Where $\hat{\sigma}_f^2$ is the moment estimate for the variance of the forward data $\{x_{t-L}, \cdots, x_{t-2}\}$, and $\hat{\sigma}_b^2$ is the moment estimate for the variance of the backward data $\{x_{t+2}, \cdots, x_{t+L}\}$. Dividing by the estimated variance is to make the scores more adaptive to the data stream with varying variance.

Predictions based on local polynomial fitting do not need pre-selected parametric models, and can be adjusted to both the linear and nonlinear data streams. Furthermore, the window bandwidth h is always small enough to keep the mined outliers outside the local data, which otherwise may degrade the detection performance in parametric methods. So we believe that our method is simpler and effective, and more convenient for implementation.

4.2 Fuzzy Partition and Decision

According to the definition of outlier and change point (see Section 2), an outlier always has both large forward and backward scores, while a change point usually has a large forward score and a small backward score. Here, these characters will be used basing on fuzzy partition and decision theory to distinguish between outliers and change points.

We consider the data set $X \triangleq \{x_t\}$ as a domain, and define four fuzzy sets on it:

$$\begin{aligned} FNormalX &= \{(x_t, \mu FNormalX = S_f(x_t)) \mid x_t \in X\}, \\ BNormalX &= \{(x_t, \mu BNormalX = S_b(x_t)) \mid x_t \in X\}, \\ NotFnormalX &= \{(x_t, \mu NotFnormalX = 1 - S_f(x_t)) \mid x_t \in X\}, \\ NotBnormalX &= \{(x_t, \mu NotBnormalX = 1 - S_b(x_t)) \mid x_t \in X\}, \end{aligned} \tag{15}$$

where $S_f(x_t) \triangleq S(Score_f(x_t))$, $S_b(x_t) \triangleq S(Score_b(x_t))$, and

$$S(x) = \begin{cases} 1, & x \le a \\ 1 - 2\left(\dfrac{x-a}{b-a}\right)^2, & a < x \le (a+b)/2 \\ 2\left(\dfrac{b-x}{b-a}\right)^2, & (a+b)/2 < x \le b \\ 0, & x > b \end{cases}. \tag{16}$$

The parameters a, b in (16) are two predefined constants that are used to control the value of the membership functions.

Then, we define two fuzzy sets named as *Outlier* and *Change* respectively as

$$\begin{aligned} Outlier &= NotFnormalX \cap NotBnormalX, \\ Change &= NotFnormalX \cap BNormalX. \end{aligned} \tag{17}$$

Their membership functions are

$$\begin{aligned} \mu Outlier &= \min(\mu NotFnormalX, \mu NotBnormalX), \\ \mu Change &= \min(\mu NotFnormalX, \mu BNormalX). \end{aligned} \tag{18}$$

Finally, point x_t with high value of $\mu Outlier$ is highly probably an outlier, while point x_t with high value of $\mu Change$ is highly probably a change point.

Note that there is another character of a change point x_t. That is x_{t-1} often has a small forward score and a large backward score. Hence, if one wants to reduce the false alarm rate, he can add another four fuzzy sets:

$$\begin{aligned} PFNormalX &= \{(x_t, \mu PFNormalX = S_f(x_{t-1})) \mid x_t \in X\}, \\ PBNormalX &= \{(x_t, \mu PBNormalX = S_b(x_{t-1})) \mid x_t \in X\}, \\ NotPFnormalX &= \{(x_t, \mu NotPFnormalX = 1 - S_f(x_{t-1})) \mid x_t \in X\}, \\ NotPBnormalX &= \{(x_t, \mu NotPBnormalX = 1 - S_b(x_{t-1})) \mid x_t \in X\}, \end{aligned} \tag{19}$$

Then the data set *Change* can be revised to

$$Change = PFNormalX \cap NotPBnormalX \cap NotFnormalX \cap BNormalX, \tag{20}$$

and its membership function is

$$\mu Change = \min(\mu PFNormalX, \mu NotPBnormalX, \mu NotFnormalX, \mu BNormalX). \tag{21}$$

The possibility of being an outlier or a change can be visualized by the values of the membership functions. Analysts can set a threshold to alarm possible outliers and changes. Users can also make their own decisions according to the membership functions and the practical experience. So we believe that our method which synthesizes both statistical and fuzzy approaches will be more effective in interactive online mining of outliers and changes.

4.3 Parameter Selection

In the proposed detection method, some parameters are essential to the detection performance, such as the bandwidth *h* of the weight function, and the fitting order *p*.

It is shown in [6] that, for all choices of *p*, the optimal kernel function is *Epanechnikov kernel* which is $K(z) = \frac{3}{4}(1 - z^2)_+$. Nevertheless, some other kernels have comparable efficiency for practical use of *p*. Hence, the choice of the kernel function is not critical.

Selection of the bandwidth *h* is important for the detection performance. Too large bandwidth will result in large estimated bias, while too small bandwidth will results in large estimated variance. A basic idea for searching the optimal bandwidth is to minimize the estimated mean integrated square error (MISE) which is defined as

$$h_{opt} = \arg\min_h \int \{(Bias(\hat{m}(x)))^2 + Var(\hat{m}(x))\}dx \tag{22}$$

However, the solution of (22) is too complex for practical use. In this paper, we employ a more convenient method to find a suboptimal bandwidth. First, we set an acceptable threshold of the MISE denoted by δ and an initial value of *h*, which is $h = [X_{\max}^f - X_{\min}^f)]/(L-1)$ for forward parameter estimation, and $h = [X_{\max}^b - X_{\min}^b)]/(L-1)$ for backward parameter estimation. Here, $X_{\max}^f \triangleq \max(x_{t-L}, \cdots, x_{t-2})$, and $X_{\min}^f \triangleq \min(x_{t-L}, \cdots, x_{t-2})$. The $X_{\max}^b$ and $X_{\min}^b$ are defined similarly. If $MISE(h) > \delta$, then multiply *h* by an expanding factor $C > 1$, i.e. $h = Ch$, until it satisfies $MISE(h) \le \delta$. An advisable value of *C* is 1.1. This searching algorithm can find a reasonable *h* quickly.

From the analysis in [6], we know that local polynomial fitting with odd order is better than that with even order. Increasing fitting order will increase computational complexity. So we set $p = 1$ for most cases and add it to 3 if necessary.

5 Simulations

We evaluate our methods by numerical simulations using different data sets.

Case (1). The first data set is generated from an AR(2) model:

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + e_t, \tag{23}$$

where, e_t is a Gaussian random variable with mean 0 and variance 1, and $a_1 = 0.6, a_2 = -0.5$. The data length is 10000. The mean of data changes at time $t = 1000\Delta\tau + 1$ $(\Delta\tau = 1, 2, \cdots, 9)$ with change size $\Delta x = 10 - \Delta\tau$. Outliers occur at time $t = 1000\Delta\tau + 501$ $(\Delta\tau = 0, 1, \cdots, 9)$ with deviation size $\Delta x = 10 - 0.8(\Delta\tau + 1)$. Fig.1 (a) shows the data set 1 and the membership functions of the fuzzy sets *Outlier* and *Change* at different time points. Here, we set $a = 8,\ b = 30$. As shown in the figure, the outliers and changes can be distinguished and detected simultaneously if the size is not very small.

Fig.1 (b) shows false alarm rate versus effective alarm rate of the outlier detection for data set 1. The effective alarm points are defined as the points during the area $[t^* - 10,\ t^* + 10]$ where t^* is the true non-normal point. Three different detection methods are compared. They are the proposed method, the CF method proposed in [4], and the parametric method proposed in [5] which is denoted by CIS method. We test the outlier of size 2.8 at time *t*=8501 for 1000 independent runs. It is observed that for the linear data stream with changing mean and constant variance, the proposed method performs comparably to the other two parametric methods.

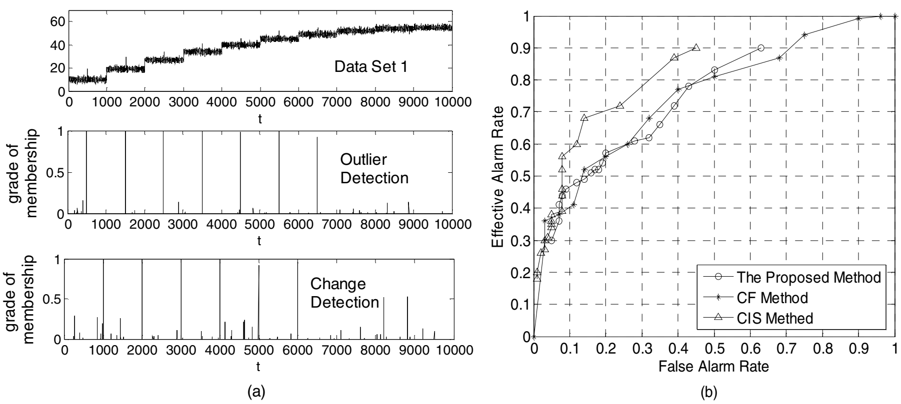

Fig. 1. Outlier and change detection for data set 1. (a) shows the data set 1, and the membership functions of *Outlier* and *Change*. (b) shows the false alarm rate vs. the effective alarm rate of outlier detection for data set 1.

Case (2). In this case, we use the similar AR(2) model as data set 1. The only difference is the variance of e_t varies gradually: $\sigma_e^2(t) = 0.1/[0.01 + (10000 - t)/10000]$. Changes and outliers occur at the same time points as data set 1, but all with size 1. The second data set and the membership functions of *Outlier* and *Change* are given in Fig.2 (a). Here we set $a = 25,\ b = 60$. Similar as the case (1), Fig.2 (b) shows false alarm rate versus effective alarm rate of the change detection for data set 2. Here, we testing the change point of size 5 at time *t*=5001. Comparing Fig.1 and Fig.2, we can see the advantage of the proposed score functions. Because of dividing by the estimated variance, the influence of the slow varying variance has been decreased a lot. That's why the proposed method outperforms the other two parametric methods in this case.

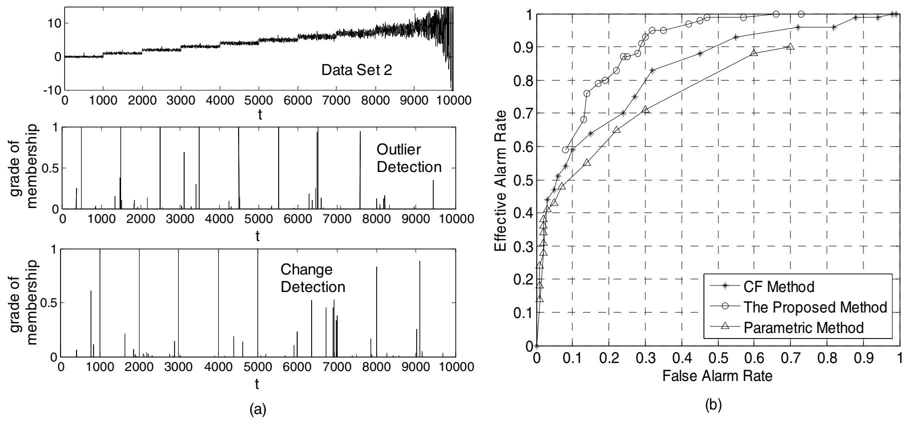

Fig. 2. Outlier and change detection for data set 2. (a) shows the data set 2, and the membership functions of *Outlier* and *Change*. (b) shows the false alarm rate vs. the effective alarm rate of change detection for data set 2.

Case (3). In this case, we change the AR(2) model to a nonlinear time series model, the ARCH(1) model:

$$X_t = \sigma_t e_t, \text{ and } \sigma_t^2 = c_0 + b_1 X_{t-1}^2. \tag{24}$$

where $e_t \sim N(0,1)$, $c_0 = 0.5$ and $b_1 = 0.5$. The mean of data also changes at $t = 1000\Delta\tau + 1$ $(\Delta\tau = 1, 2, \cdots, 9)$ with size $\Delta x = 10 - \Delta\tau$. Outliers occur at time $t = 1000\Delta\tau + 501$ $(\Delta\tau = 0, 1, \cdots, 9)$ with deviation size 7. Fig.3 (a) shows the third data set and the membership functions of *Outlier* and *Change*. Curves of false alarm rate versus effective alarm rate of outlier and change detection for data set 3 are shown in Fig.3 (b). Here, we test the outlier of size 7 at time t=1501, and the change

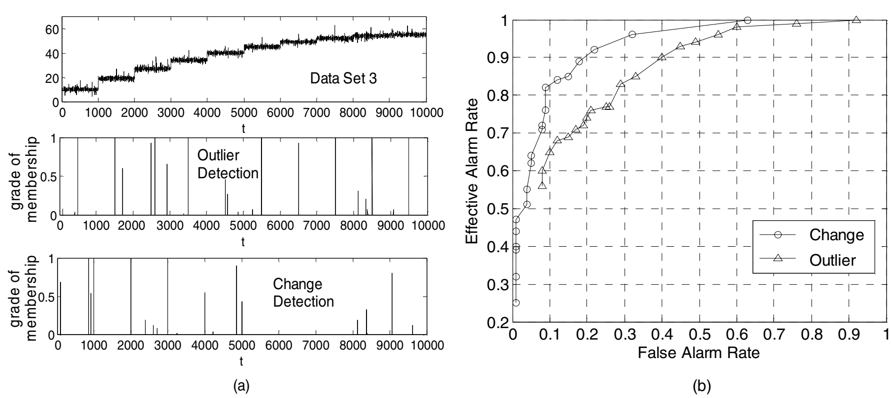

Fig. 3. Outlier and change detection for data set 3. (a) shows the data set 3, and the membership functions of *Outlier* and *Change*. (b) shows the false alarm rate vs. the effective alarm rate of outlier and change detection.

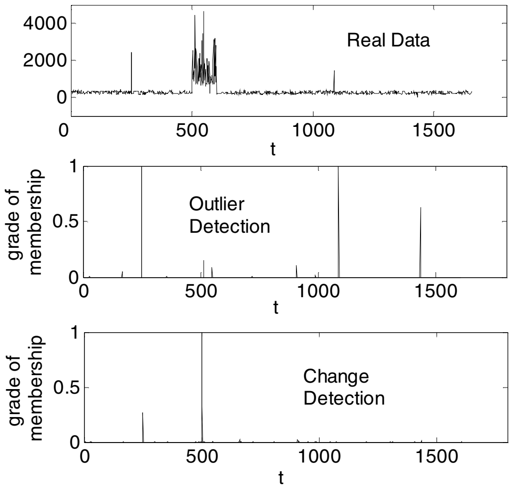

Fig. 4. Outlier and change detection for real data

point of size 3 at time t=7001. It is easy to see the proposed nonparametric detection method is also appropriate to the nonlinear data streams, which is difficult for the parametric methods.

Case (4). The real data case. Here, we test our method by a real data set sampling from the dataset KDD Cup 1999 which is prepared for network intrusion detection. There are 3 intrusions in this real data set, respectively at time t=250, t=1087, and t=1434. The mean and variance of the normal data suddenly change at t=501, and recover at t=602. We present the real data set and the membership functions of *Outlier* and *Change* in Fig.4. It is shown that the proposed method is effective in the real data case. The intrusions are detected as outliers, and the sudden change of the normal data is detected as change points.

6 Conclusion

This paper presents a unifying method for outlier and change detection from data streams. Unlike conventional parametric methods, the proposed method is based on a nonparametric technique, the local polynomial fitting. Fuzzy partition and decision method are used to alarm possible outliers and changes. The proposed method is more appropriate to online and interactive data mining. Simulation results reveal its robustness and efficiency.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (No.10571127) and the Specialized Research Fund for the Doctoral Program of Higher Education (No.20040610004).

References

1. G. Dong, J. Han, L.V.S. Lakshmanan, J. Pei, H. Wang and P.S. Yu: Online Mining of Changes from Data Streams: Research Problems and Preliminary Results. Proceedings of the ACM SIGMOD Workshop on Management and Processing of Data Streams. ACM, New York (2003) 225-236
2. Zakia Ferdousi and Akira Maeda: Unsupervised Outlier Detection in Time Series Data. Proceedings of ICDEW'06. IEEE Computer Society, Washington, DC (2006) 51-56
3. M. Garofalakis, J. Gehrke, and R. Rastogi.: Querying and mining data streams: You only get one look. Proceedings of SIGMOD'02. ACM, New York (2002) 635-642
4. Jun-ichi Takeuchi and Kenji Yamanishi: A Unifying Framework for Detecting Outliers and Change Points from Time Series. IEEE Transaction on Knowledge and Engineering. IEEE Press, USA (2006) 482-492
5. Zhi Li, Hong Ma, Yongdao Zhou: A Unifying Method for Outier and Change Detection from Data Streams. Proceedings of CIS'06. IEEE Press, USA (2006) 580-585
6. Jianqing Fan and Qiwei Yao: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York (2002) 215-246.
7. H. D. Cheng, Y. H. Chen and X. H. Jiang: Unsupervised Change Detection Using Fuzzy Entropy Principle. IEEE International Geoscience and Remote Sensing Symposium. IEEE Press, USA (2004) 2550-2553