

Outlier Detection in Large-Scale Traffic Data by Naïve Bayes Method and Gaussian Mixture Model Method

Philip Lam¹, Lili Wang², Henry Y.T. Ngan¹, Nelson H.C. Yung³, Anthony G.O. Yeh⁴

¹Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

²Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong

³Department of Electronic & Electrical Engineering, ⁴Department of Urban Planning & Design, The University of Hong Kong, Hong Kong

Abstract

It is meaningful to detect outliers in traffic data for traffic management. However, this is a massive task for people from large-scale database to distinguish outliers. In this paper, we present two methods: Kernel Smoothing Naïve Bayes (NB) method and Gaussian Mixture Model (GMM) method to automatically detect any hardware errors as well as abnormal traffic events in traffic data collected at a four-arm junction in Hong Kong. Traffic data was recorded in a video format, and converted to spatial-temporal (ST) traffic signals by statistics. The ST signals are then projected to a two-dimensional (2D) (x, y) -coordinate plane by Principal Component Analysis (PCA) for dimension reduction. We assume that inlier data are normal distributed. As such, the NB and GMM methods are successfully applied in OD (Outlier Detection) for traffic data. The kernel smooth NB method assumes the existence of kernel distributions in traffic data and uses Bayes' Theorem to perform OD. In contrast, the GMM method believes the traffic data is formed by the mixture of Gaussian distributions and exploits their respective confidence region for OD. This paper would address the modeling of each method and evaluate their respective performances. Experimental results show that the NB algorithm with Triangle kernel and GMM method achieve up to 93.78% and 94.50% accuracies, respectively.

Introduction

It is desired that most events to be safe, stable or rather predictable. Identifying an unusual event is a typical and vital topic in different fields, such as intrusion detection in cyber security, fraud detection for credit cards, insurance or health care and fault detection in safety critical systems [9]. We may call this unusual event as outlier or abnormality which is different from a usual event. Outlier is usually minor in a group of events/data while inlier is the majority.

OD refers to detect any abnormal element in data which is not consistent with an expected behavior [5]. A good OD method should be accurate to detect outlier and less erroneous judgment on inlier data. Many OD methods [8] have been developed in recent years. In [5, 9], nearest neighbor approach was proposed to consider the distance or the similarity between two data instances. The assumption behind the approach is that inliers in data should be dense and outlier(s) is far from these dense inliers.

NB classifiers were developed in the 1950s. The NB classifiers are constructed based on Bayes' theorem, and are widely used for text categorization with superior performance [15]. Another popular application is spam filtering among numerical emails [16]. The kernel smooth NB method in this paper supposes the traffic data could be modeled by kernel distributions so that the Bayes' theorem

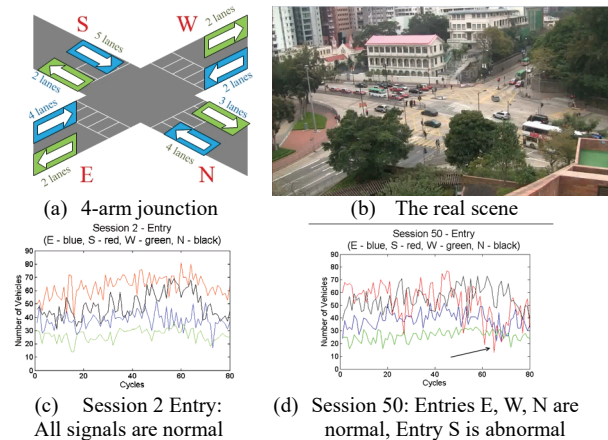


Figure 1. (a) A generic diagram of the 4-arm junction; (b) sample of the real scene; (c) normal ST signals; (d) abnormal ST signals.

is applied. In contrast, the GMM method presumed a mixture of Gaussian distributions and confidence region is employed. GMM, firstly introduced in [2], is an appropriate way of handle data with multiple outliers [17]. The GMM method is useful in dealing with speech recognition [18].

For traffic data, detecting anomaly traffic event would be better to deal with the traffic problems. This research aims at detecting outliers for a large-scale traffic database from Hong Kong. The original video was taken at a four-arm junction as shown in Fig. 1(a). The 4-arm junction can be expressed as an ideal map like Fig. 1(b). In total, the traffic data for 31 days was recorded with two sessions per day: AM (07:00-10:00) and PM (17:00-20:00). For each session, the original video data are dividing into 19 spatial temporal (ST) signals. Fig. 1(c) demonstrates a sample of 4 normal ST signals. These ST signals have different number of traffic cycles because smoother traffic flow would lead to a shorter traffic cycle and result in more traffic cycles in one session. Also, the ST signals in each direction suffer a high degree of similarities (Fig. 1(d)) among each other or across signals in different traffic direction. Therefore, it requires a truncation process (to standardize the length of a signal's cycles) and a signal domain transformation (to remedy the signal similarities). Herein, each ST signal was processed by principal component analysis (PCA) [11] to reduce the dimension of signal representation. Then, an OD method was carried out based on the PCA-processed (x, y) -coordinates plane.

In this paper, a detail investigation of NB and GMM classifiers in large-scale traffic data is carried out, for these two methods are newly applied in traffic OD. Experimental results demonstrate that

the NB and GMM methods can achieve 93.78% and 94.50% detection success rates (DSRs), respectively. These performance is comparable to previous evaluation on other OD methods [12, 13] including Gaussian mixture model (80.86%), one-class SVM (59.27%), S-estimator (76.20%) and kernel density estimation (95.20%).

The organization of this paper is as follows: Section II gives a review of related work about OD. Sections III and IV present details of two proposed OD methods and their experimental results, and Section V concludes this paper.

Related Work

The popular OD methods include statistical, nearest neighborhood, spectral approaches and learning based approaches. Details about these methods are given as follows.

Statistical Approach

Statistical approach is one of the earliest approaches in OD. It assumes that all inliers occur in a high probability region of this distribution model when outliers deviate strongly from the inliers. Statistical OD methods can perform parametric techniques like GMM or regression, or nonparametric techniques like histogram based or kernel function based. This approach would perform well if data distribution is assumed well. However, if the data distribution assumption is false, the result would be far away from the correct situation [3, 8].

Nearest Neighbor Approach

Nearest neighbor approach considers the distance or the similarity between two data instances. The assumption behind the approach is that inliers in data should be dense and outlier(s) is far from these dense inliers [9, 5]. This approach can be easily realized in an unsupervised way that could require high computational complexity. As a result, it is hard to deal with very complex data [8].

There are two major variant methods related to this approach, density measurement based methods, such as Local Outlier Factor (LOF) [19], Influenced Outlierness (INFLO) [3] and Local Outlier Correlation Integral (LOCI) [3], or distance measurement based methods, like Mahalanobis Distance [6], $DB(\epsilon, \pi)$ -Outliers, Outlier Scoring based on k-NN distances [4], Resolution-based Outlier Factor (ROF) [8].

Spectral Approach

Spectral approach suggests that inliers and outliers could appear significantly different in a spectral domain. The approach detects outliers by embedding data into a lower dimensional subspace. The approach is useless if inliers and outliers in data are not separable in the lower dimensional subspace. In addition, it requires high computational complexity [8].

Learning Approach

Learning approach is using various training methodologies [10] (or so-called trained machines) to train the input data. The related OD methods are neural networks, naïve Bayesian network, support vector machines (SVM). The assumption is that inliers and outliers in data can be easily distinguished by the trained machines. Therefore, the machines can test and classify a test instance into either an inlier group or an outlier group in OD [8]. There are many algorithms that can be used in multi-learning class approach, and they have a faster testing phase than other approach. However, accurate labeling for various normal classes is often impossible [8].

In this paper, an investigation of NB classifier and GMM classifier to model the PCA-processed (x,y)-data and their performance for OD in traffic data would be carried out.

Kernel Smoothing NB Method

NB classifier was developed in 1950s for text retrieval [18]. It is also widely used for spam filtering among numerical emails [1].

a. NB Classifier

NB classifier [3] is developed based on Bayes' Theorem which takes the form of

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} \quad (1)$$

Based on the probability $P(E)$, $P(H)$, and the conditional probability $P(E|H)$, the posteriori probability $P(H|E)$, which denotes the possibility of event H conditioned on an occurred event E, can be obtained. Naive Bayes classifiers are based on the information of the training data, and then determined the highest possible class of testing data from their information.

b. Kernel Distribution

Kernel Distribution is a non-parametric distribution to estimate each training point as some independent distributions into the whole distribution [7].

$$\hat{f}_h = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \quad (2)$$

Kernel Distribution \hat{f}_h is estimated by n sample data points with a kernel density $K_h(x - x_i)$. The commonly used kernels are listed as follows:

Box kernel:

$$K(x) = 0.5I\{|x| \leq 1\} \quad (3)$$

Triangle kernel:

$$K(x) = (1 - |x|)I\{|x| \leq 1\} \quad (4)$$

Epanechnikov kernel:

$$K(x) = 0.75(1 - x^2)I\{|x| \leq 1\} \quad (5)$$

Normal kernel:

$$K(x) = \frac{1}{2\pi} \exp(-0.5x^2) \quad (6)$$

where x is replaced by y for y axis kernel.

c. Kernel Smoothing NB classifier for OD

Kernel smoothing NB (KSNB) classifier is a trained Kernel Distribution based classifier. Different kernels with different widths for each predictor or class are available in KSNB classifier. We use the trained KSNB classifiers to detect outliers, where the classifiers would automatically set a bandwidth value for each feature and class which is optimal for a Gaussian model. Then, a corresponding region from contours surrounding various kernel distributions is formed an inlier region E, therefore any data out of that region is classified as an outlier. In this method, box, triangle, Epanechnikov and normal kernels are used for OD. The OD procedures based on KSNB classifier are given as follows:

d. Procedure

Fig. 2 illustrates the flowchart of the proposed NB based method (See Algorithm 1), which includes a training stage and a testing stage. The details are as below.

(i) Training stage

Step 1: Feed training data (all inliers) into NB classifiers.

Step 2: Fit training data into the specified kernel distributions defined in Eqns. (3-6).

Step 3: Combine each kernel distribution into a whole distribution.

Step 4: Set a bandwidth (BIB) of the inlier region automatically from the kernel distribution.

Algorithm 1. NB classifier with different kernel distribution

Require: The inliers data set $A = (x_i, y_i)$ and testing data set $B = (x_i, y_i)$

- 1: define outliers set H
- 2: **for** every data point A
- 3: form in kernel distribution $(K_i(x - x_i), K_i(y - y_i))$
- 4: **end**
- 5: $\hat{f}_h = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i)$
- 6: set the bandwidth value by \hat{f}_h
- 7: form an inlier region E
- 8: **for** all i in B
- 9: **if** $(x_i, y_i) \in E$
- 10: $(x_i, y_i) \in \sim H$
- 11: **Else** $(x_i, y_i) \in H$
- 12: **end**
- 13: output labels

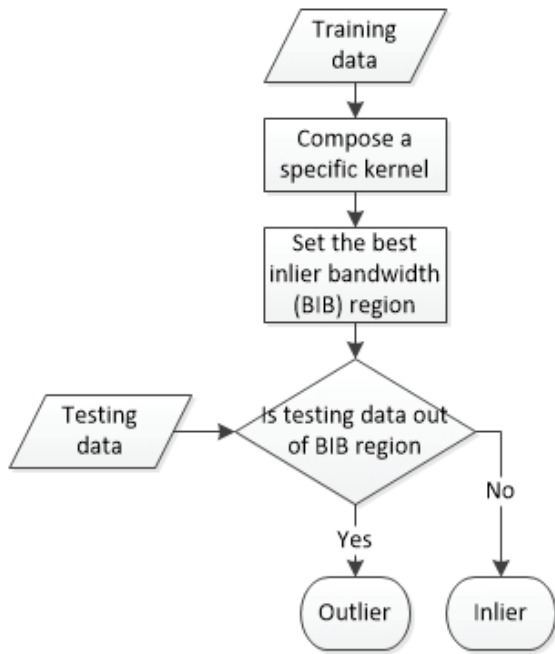


Figure 2. Flowchart of the proposed NB method.

(ii) Testing Stage

- Step 1: Input testing data into the trained NB classifiers.
- Step 2: Data label as outlier if the data is out of the inlier region.
- Step 3: Output result.

e. Experimental results

Experimental results based on the kernel smoothing NB method are listed in Table I. Accuracy (Acc.), positive predictive value (PPV), negative predictive value (NPV), sensitivity (Sen.), false positive rate (FPR) are employed as measurement metrics and their definitions can be referred to [12]. From the results, we can see

that overall accuracy among all kernels is almost higher than 90%. Overall NPV among all kernels is higher than 95%.

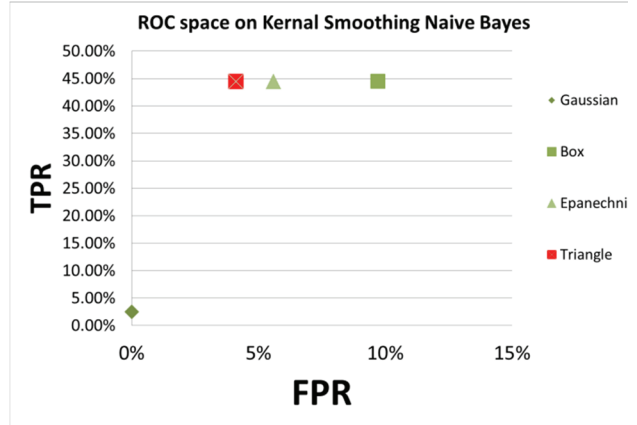


Figure 3. ROC plot for the NB method based on four kernels.

TABLE I. PERFORMANCE OF THE NB METHOD AMONG DIFFERENT KERNELS.

	Kernel	Gaussian	Box	Epanechnikov	Triangle
AM	Acc.	97.13	91.39	95.22	96.17
	PPV	NA	18.18	40.00	50.00
	NPV	97.13	97.93	98.09	98.09
	Sen.	0.00	33.33%	33.33	33.33
	FPR	0.00	6.84	2.87	1.91
PM	Acc.	91.87	85.65	89.47	91.39
	PPV	100.00	22.71	29.49	36.81
	NPV	91.65	96.21	96.49	96.49
	Sen.	5.00	55.50	55.50	55.50
	FPR	0.00	12.60	8.31	6.30
All	Acc.	94.50	88.52	92.345	93.78
	PPV	100.00	20.45	34.75	43.41
	NPV	94.39	97.07	97.29	97.29
	Sen.	2.50	44.42	44.42	44.42
	FPR	0.00	9.72	5.59	4.11

The OD scatters corresponding to the four kernels are shown in Fig. 3. From the figure, one observes that the scatter based on Triangle kernel is closest to the perfect classification point. Among the NB method with different kernels, Gaussian kernel usually has better performance. However, in our evaluation, the OD performance based on NB method with Gaussian kernel is very close to the random-guess line ($y = x$ black line) as shown in ROC space of Fig. 3. It is just a little better than random guessing that all testing data are inliers. Therefore, Gaussian kernel is found to be not a good kernel in this method for OD. For other three kernels listed in Table I, the triangle kernel shows the best performance that is closest to the perfect classification point with the highest accuracy (93.78%). For AM or PM OD, the NB method with the triangle kernel still outperforms other three kernels That may indicate that the NB method with a triangle kernel is significantly good to describe the behaviors of PCA traffic data.

GMM Method

The GMM method was firstly introduced by Aitkin and Wilson in 1980 [2], the mixture model as a way of handle data with multiple outliers [4]. The GMM method is good in deal with speech recognition [18].

a. Gaussian Distribution

In mathematics, a one-dimensional **Gaussian distribution** is a distribution function of the possibility density form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \dots e^{-\frac{1}{2}\left(\frac{x-\mathcal{M}}{\sigma}\right)^2} \quad (7)$$

where $x \sim N(\mathcal{M}, \sigma^2)$, \mathcal{M} denotes the expectation of x , σ^2 denotes the variance of x

In term of a two-dimensional expression, the Gaussian distribution is

$$f(x, y) = e^{-\frac{1}{2}\left(\frac{(x-\mathcal{M}_x)}{\sigma_x} + \frac{(y-\mathcal{M}_y)}{\sigma_y}\right)^2} \quad (8)$$

where $x \sim N(\mathcal{M}_x, \sigma_x^2)$ and $y \sim N(\mathcal{M}_y, \sigma_y^2)$.

b. GMM

GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. The p.d.f. of the mixture model among f_i is $F(x, y)$ which takes the form of

$$F(x_i, y_i) = \sum_{i=1}^n w_i f_i(x_i, y_i) \quad (9)$$

where w_i is the mixture weight [14].

c. Methodology

For OD in this section, we will use confidence region (CR) with a Bonferroni adjustment as criterion in α significant level. Herein, $\alpha = 0.1, 0.05, 0.01$.

For a CR with Bonferroni adjustment, GMM CR for X is

$$[\bar{x} - t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(x)}, \bar{x} + t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(x)}] \quad (10)$$

GMM confidence region for Y is

$$[\bar{y} - t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(y)}, \bar{y} + t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(y)}] \quad (11)$$

where t_k is student's value in k significant level. The GMM method models fit the training data by using several Gaussian distributions.

In this section, data fitting will use the expectation maximization (EM) algorithm [14], which assigns posterior probabilities to each component density with respect to each observation, to get the best w_i for each i in (9).

d. Procedure

A flowchart of the proposed GMM method, including a training stage and a testing stage is presented in Fig. 4 (Also see Algorithm 2). The details are as below.

(i) Training stage

Step 1: Input inlier training data.

Step 2: Fit each data into individual identical Gaussian distribution model.

Step 3: Mixture all models into a whole GMM model (9).

Step 4: Set the criteria region for different criterion method and different significance level [(i) 0.10, (ii) 0.05, (iii) 0.01].

(ii) Testing Stage

Step 1: Input testing data set.

Step 2: Label data as an outlier if the data is out of the criteria region.

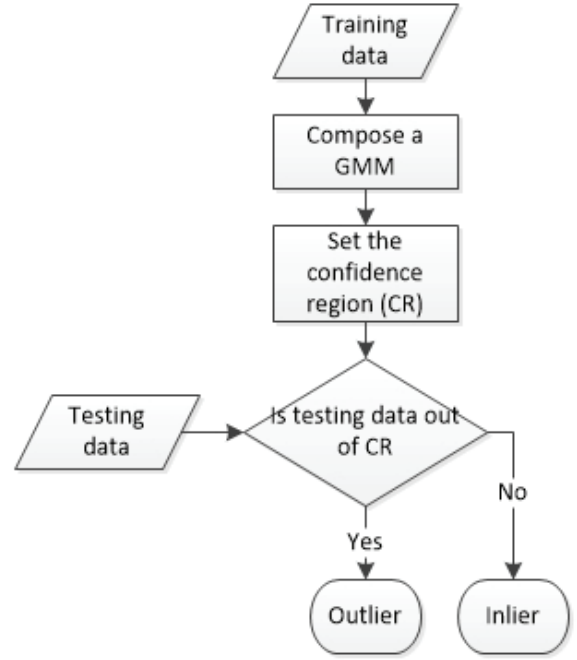


Figure 4. The flowchart of the proposed GMM method.

Algorithm 2. GMM with 1 mixture classifier in confidence region method

Require: The inliers data set $A = (x_i, y_i)$ and testing data set $B = (x_i, y_i)$

- 1: define outliers set H
 - 2: **for** every data point A
 - 3: form a Gaussian model with
 - 4: $GMM \sim N\left((\bar{x}, \bar{y})^T, \begin{pmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(x, y) & \text{var}(y) \end{pmatrix}\right)$
 - 5: **end**
 - 6: fixed the significant level α
 - 7: construct **GMM x** confidence region X with:
 - 8: $[\bar{x} - t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(x)}, \bar{x} + t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(x)}]$
 - 9: construct **GMM y** confidence region Y with:
 - 10: $[\bar{y} - t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(y)}, \bar{y} + t_{\frac{\alpha}{2+3}} \sqrt{\text{var}(y)}]$
 - 11: **for** all i in B
 - 12: **if** $x_i \in X$
 - 13: **if** $y_i \in Y$
 - 14: $(x_i, y_i) \in \sim H$
 - 15: **else** $(x_i, y_i) \in H$
 - 16: **end**
 - 17: **else** $(x_i, y_i) \in H$
 - 18: **end**
 - 19: **end**
 - 20: Output results
-

e. Experimental results

Experimental results based on the GMM method are listed in Table II. The performances in terms of FPR, PPV and Acc. are gradually better with the decrease of significant level. The highest accuracies in both AM (96.65%) and PM (94.50%) sessions are achieved when the significant level is 0.01. A ROC plot on the GMM method is shown in Fig. 5. By observation, CR in the 0.01 level is the closest point to the perfect classification point. Inlier data points lying in the centroid of the GMM model and the CR cover at least α significant level built by training data set. Therefore, the CR criterion is good at describing the inlier group which is centered.

TABLE II.
PERFORMANCE OF THE GMM METHOD IN 3 SIGNIFICANT LEVELS.

Level		0.10	0.05	0.01
AM	Acc.	89.47%	93.78%	96.65%
	PPV	9.09%	14.29%	50.00%
	NPV	97.70%	97.99%	98.09%
	Sen.	33.33%	33.33%	33.33%
	FPR	9.04%	4.50%	1.44%
PM	Acc.	82.30%	86.12%	92.34%
	PPV	18.83%	28.02%	46.30%
	NPV	96.26%	96.52%	96.11%
	Sen.	55.50%	55.50%	53.50%
	FPR	16.55%	11.81%	4.55%
Overall	Acc.	85.89%	89.95%	94.50%
	PPV	13.96%	21.16%	48.15%
	NPV	96.98%	97.26%	97.10%
	Sen.	44.42%	44.42%	43.42%
	FPR	12.80%	8.16%	3.00%

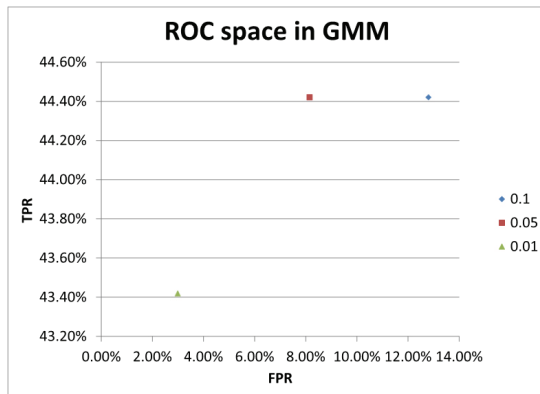


Figure 5. ROC plot for the GMM method based on different significant levels.

Conclusion

In this paper, we present two OD methods, kernel smoothing NB and GMM, to detect outliers in large-scale traffic data. The kernel smoothing NB method utilizes the trained kernel smoothing NB classifiers to detect outliers, in which the classifiers set the best bandwidth value with the inlier region. Any data points out of that region are classified as outliers. In the GMM method, the rectangular confidence region is formed in Bonferroni adjustment way. True α significant level region of the GMM can be constructed as the inlier region for the more accuracy. Experimental results show that the two

algorithms can achieve pleasing detection accuracies compared with the OD methods in our previous studies [12, 13], including Gaussian mixture model, one-class SVM, S-estimator and kernel density estimation.

Acknowledgement

This research is supported by Hong Kong RGC GRF: 12201814, HKBU FRG2/14-15/075, HKBU FRG1/15-16/002.

References

- [1] C. Aggarwal, *Outlier Analysis*, Springer, 2013.
- [2] M. Aitkin and G.T. Wilson, "Mixture models, outliers, and the EM algorithm," *Technometrics*, 22, 325-331, 1980.
- [3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, 41(3), pp.1-58, 2009.
- [4] V. Hautamaki, I. Karkkainen and P. Franti, "Outlier Detection Using k-Nearest Neighbour Graph," *Proc. IEEE ICPR*, vol. 3, 430-433, 2004.
- [5] V. J. Hodge, J. Austin, "A Survey of Outlier Detection Methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp.85-126, 2004.
- [6] G.S.D.S. Jayakumar and B. J. Thomas "A New Procedure of Clustering Based on Multivariate Outlier Detection", *Science*, 11, 69-84, 2013.
- [7] G.H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," *In: Hanks, P. B. eds., Proc. 11th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, CA, pp. 338-345, 1995.
- [8] H. Kriegel, P. Kroger and A. Zimek, "Outlier Detection Techniques," *Tutorial at the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2009.
- [9] V. Kumar, "Parallel and Distributed Computing of Cybersecurity," *IEEE Distributed Systems Online*, vol. 6, no. 10, 2005.
- [10] M. Minsky, S. Papert, "An Introduction to Computational Geometry," *MIT Press*, ISBN 0-262-63022-2, 1969.
- [11] H.Y.T. Ngan, N.H.C. Yung and A.G.O. Yeh, "Detection of Outliers in Traffic Data based on Dirichlet Process Mixture Model," *Proc. IEEE CASE*, pp 224-229, 2012.
- [12] H.Y.T. Ngan, N.H.C. Yung and A.G.O. Yeh, "Detection of Outliers in Traffic Data based on Dirichlet Process Mixture Model," *IET Intelligent Transportation Systems*, vol. 9, no. 7, pp. 773-781, 2015.
- [13] H.Y.T. Ngan, N.H.C. Yung and A.G.O. Yeh, "A Comparative Study of Outlier Detection for Large-scale Traffic Data by One-class SVM and Kernel Density Estimation," *IS&T/SPIE Electronic Imaging*, 94050I-10, 2015.
- [14] D.A. Reynolds, "Gaussian Mixture Models," *Encyclopedia of Biometric Recognition*, Springer, Heidelberg, 2008.
- [15] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed., Prentice Hall, ISBN 978-0137903955, 2003.
- [16] M. Sahami, S. Dumais, D. Heckerman and E. Horvitz, "A Bayesian approach to filtering junk e-mail," *AAAI Workshop on Learning for Text Categorization*, AAAI Technical Report, WS-98-05, 1998.
- [17] D.W. Scott, "Outlier detection and clustering by partial mixture modelling," *COMPSTAT Symposium*, pp. 453-465, Heidelberg, 2004.
- [18] M. N. Stuttle, "A Gaussian mixture model spectral representation for speech recognition," *Ph.D. dissertation, University of Cambridge*, 2003.
- [19] K-A. Yoon, O-S. Kwon, and D-H. Bae, "An Approach to Outlier Detection of Software Measurement Data Using the K-means Clustering Method," *IEEE 1st Int'l Sym. ESEM*, pp. 443-445, 2007.

Author Biography

Philip Lam received the B.Sc. (Hons) degree in Statistics and Operations Research (2015) in Hong Kong Baptist University, China. He is studying the M.Sc. in Actuarial and Investment Science at Hong Kong Polytechnic University, China and is supposed to obtain the degree in 2017. Also, he is currently an actuarial intern in China Taiping Life Insurance (Hong Kong) Company Limited.

Li-Li Wang received the M.Sc. degree from Shandong University, Shandong, China (2007) and the Ph.D. degree from the Hong Kong Polytechnic University, China (2013). She is currently a postdoctoral fellow in the Department of Health Technology and Informatics, the Hong Kong Polytechnic University. Her research interests include digital signal processing, image and video coding, pattern recognition, machine learning, computer vision and big data analysis.

Henry Y.T. Ngan received the B.Sc. degree in Mathematics (2001), the M. Phil. degree (2005) and the Ph.D. degree (2008) in Electrical & Electronic Engineering at The University of Hong Kong, China. He is currently an assistant professor of research in Mathematics, Hong Kong Baptist University. He was a conference chair of IS&T Electronic Imaging 2016, 2017. He is an editor of IS&T Journal of Imaging Science & Technology.

Nelson H.C. Yung received his B.Sc. and Ph.D. degrees from Newcastle University, UK. He has co-authored five books and book chapters and has published over 190 journal and conference papers in the areas of digital image processing. He was a guest editor of the SPIE Journal of Electronic Imaging. He is a Chartered Electrical Engineer, Member of the HKIE, IET and Senior member of the IEEE.

Anthony G.O. Yeh received his BA degree in geography from the University of Hong Kong (1974), MSc from AIT, Thailand (1976), MRP (1978) and PhD (1980) in urban and regional development and planning from Syracuse University, USA. He is Chan To-Haan Professor in Urban Planning and Design in the Department of Urban Planning and Design of the University of Hong Kong and an Academician of the Chinese Academy of Sciences and a Fellow of TWAS.