2010 International Conference on Multimedia Information Networking and Security

Fault Diagnosis Analysis in Large-Scale Computing Environments

Yan Xue
Dept. of Information Management in Nanjing University
Nanjing, Jiangsu
e-mail: xueyan.nju@gmail.com

Xuefang Zhu
Dept. of Information Management in Nanjing University
Nanjing, Jiangsu
e-mail: xzf@nju.edu.com

Abstract—**This paper issues the problem of fault diagnosis in high computing system. In order to solve this problem, i.e., correctly and efficiently detecting the anomaly nodes during the system operation, which is very similar to the principle of pattern recognition research work, thus we try to use some pattern recognition methods to analysis and solve fault diagnosis problem in this paper. And also we do some experiment and compare the results and finally get some useful conclusion to show that Kernel Eigenface and Kernel Fisherface methods achieve lower error rates than the ICA and PCA approaches in anomaly nodes detection.**

Keywords-fault diagnosis; pattern recognition; feature extraciton; PCA; KPCA; ICA; FDA;KFDA

I. INTRODUCTION

As the scale of high performance computing (HPC) grows, the new generation of HPC system involves tens-of-thousands to hundreds-of-thousands of components. For systems of this scale, reliability becomes a major concern as the system wide Mean-Time-Before-Failure (MTBF) decreases dramatically with the increasing count of components. Therefore, fault diagnosis is a critical point for long-running parallel distributed applications executing in Massive Cluster of Workstations (MCOW).

In order to solve this problem, a number of failure analysis and prediction methods have been presented to detect when a system fails to function properly, such as model-based predictive methods to trigger an alert when a deviation from the derived model is observed; exploited data mining techniques to detect critical events in large-scale clusters by capturing and discovering fault patterns. While it is useful for system management by detecting when the system functions abnormally, it is more critical to find out which part of the system is the source of the problem. In other words, the key challenge therefore is how to effectively determine the source of the problem (e.g. a subset of nodes) from the potentially overwhelming amount of information collected across the system.

In this paper, we try some methods to detect the anomaly data based on [1]. By observation, [1] shows that the nodes perform comparable activities generally exhibit similar behaviors. Thus it is urgent to explore such a similarity to find a small set of nodes that are substantially different from the majority. Thus, it involves three key questions:

(1) How to collect a variety of features for effectively representing node behaviors?

(2) How to extract the most significant features for anomaly localization from the potentially overwhelming mass of data?

(3) How to quickly identify faulty nodes?

Zheng etc. in [1] tries to solve these problems by using such three steps:

(1) **feature collection** to assemble a feature space for the system (generally has high dimensionality to capture a wide variety of system features);

(2) **feature extraction** to obtain the most significant features out of the original feature space for efficient data analysis;

(3) **outlier detection** to quickly identify the nodes that are "far away" from the majority.

Zheng etc. proposed to use PCA and ICA methods in feature extraction process. But in fact PCA can be effectively performed on a set of data that vary linearly, which means that if the data are nonlinear, PCA could not work well. In addition, Independent Component Analysis (ICA) [6] is yet another linear decomposition which seeks statistically *independent* and non-Gaussian components, modeling the observed data as a linear mixture of (unknown) independent sources. In other words, when the data are non-Gaussian distributed, ICA method could perform well to category them. But once the data are Gaussian distributed, ICA would not work well. In order to solve such problems, we try to use Kernel method in feature extraction. In fact, in face

978-0-7695-4258-4/10 $26.00 © 2010 IEEE
DOI 10.1109/MINES.2010.220

recognition, some results have shown that Kernel method outperforms PCA and ICA methods. In other words, kernel methods could effectively localize anomalies in massive quantities of data.

In this paper we seek a method that not only extracts higher order statistical dependencies of samples as features, but also maximizes the class separation when we project these features to a lower dimensional space for efficient recognition. Since much of the important information may be contained in the high order dependencies among pixels of a face image, we investigate the use of Kernel PCA and Kernel FLD for anomaly detection, which we call Kernel Eigenface and Kernel Fisherface methods, and compare their performance against the standard Eigenface, and ICA-based methods.

Kernel Principal Component Analysis (KPCA) is a nonlinear extension of PCA and the essential idea is to map input space to a higher dimensional feature space, through a non-linear map where the data is linearly separable. Thus it could effectively compute the principal components in a big dimensional feature space, which is non-linearly related to the input space.

Kernel Fisher Linear Discriminate (KFLD) method is just to apply the kernel trick to FLD and experiments showed that KFLD is able to extract the most discriminate features in the feature space, which is equivalent to extracting the most discriminate nonlinear features in the original input space.

The rest of this paper is organized as follows. Section 2 provides a brief discussion of the feature extraction methods to detect the anomaly nodes in high computing system. Section 3 gave some detailed discussion about the research result based on the feature extraction method proposed in section 2 and showed that the result is better than [1]. Finally we conclude the paper in section 4.

II. ANAMOLY LOCALIZATION BASED ON PARRTERN RECOGNITION METHOD

As analysis above, how to correctly and efficiently recognize the anomaly nodes, i.e., feature extraction, is the key problem we need to solve in high performance computing system. Therefore, in order to solve this problem, we will analysis some pattern recognition methods in the following and try to use them to detect anomaly nodes and make analysis in experiment.

1 PCA

Consider a linear transformation mapping the original n-dimensional image space into an m-dimensional feature space, where m<n.

The new feature vectors $y_k \in R^m$ are defined by the following linear transformation:

$$y_k = W^T x_k \qquad (1)$$

where k=1,2,...,N and W is a matrix with orthonormal columns. Different objective functions will yield different algorithms with different properties. PCA aims to extract a subspace in which the variance is maximized. Its objective function is as follows:

$$W_{opt} = [w_1, w_2, ..., w_m] = \arg\max_W \left| W^T S_t W \right| \qquad (2)$$

Within total scatter matrix is defined as

$$S_t = \sum_{k=1}^{n} (x_k - \mu)(x_k - \mu)^T \qquad (3)$$

Where μ is the mean of all samples.

The optimal projection $W_{opt} = [w_1, w_2, ..., w_m]$ is the set of n-dimensional eigenvectors of S_t corresponding to the m largest eigenvalues, i.e.

$$S_t w_i = \lambda_i w_i,\ i = 1, 2, 3, ..., m \qquad (4)$$

2 Kernel PCA

Given a set of m centered (zero mean, unit variance) samples $x_i,\ i = 1, 2, ...N$,

x_i is projected from the input space R^m to a high dimensional feature space R^f by a nonlinear mapping function $\phi : R^n \to R^f$

In R^f, the corresponding eigenvalue problem (4) becomes

$$S_t^{\phi} w^{\phi} = \lambda w^{\phi} \qquad (5)$$

Then the scatter matrix can be re-calculated in R^f as follows:

$$S_t^{\phi} = \sum_{k=1}^{N} \phi(x_k)\phi(x_k)^T = A_t^{\phi} (A_t^{\phi})^T \qquad (6)$$

Where $A_t^{\phi} = [\phi(x_1), \phi(x_2), ..., \phi(x_N)]$ is a matrix whose column are $\phi(x_i)$.

And there exist coefficient vector $\alpha = [\alpha_1, \alpha_2, ..., \alpha_N]^T$ such that

$$w^{\phi} = \sum_{k=1}^{N} \alpha_k \phi(x_k) = A_t^{\phi} \alpha \qquad (7)$$

Denoting a N*N matrix K by

$$W_{OPT}^{\phi} = \arg\max_{W^{\phi}} \frac{\left|(W^{\phi})^T S_B^{\phi} W^{\phi}\right|}{\left|(W^{\phi})^T S_w^{\phi} W^{\phi}\right|} = [w_1^{\phi}, w_2^{\phi}, \ldots, w_m^{\phi}]$$

$$K_{ij} = k(x_i, x_j) = \phi(x_i)^T \phi(x_j) = (A_t^{\phi})^T A_t^{\phi} \quad (8)$$

Then Kernel PCA problem (5) becomes

$$A_t^{\phi} (A_t^{\phi})^T A_t^{\phi} \alpha = \lambda A_t^{\phi} \alpha$$
$$\Leftrightarrow A_t^{\phi} K\alpha = A_t^{\phi} \lambda\alpha$$
$$\Leftrightarrow K\alpha = \lambda\alpha$$

So firstly we solve the equation $K\alpha = \lambda\alpha$. (9)

Then we can project the vectors in R^f to a lower dimensional space spanned by the eigenvectors w^{ϕ} . Then the projection of x to the eigenvectors w^{ϕ} is

$$w^{\phi}\phi(x) = \sum_{k=1}^{N} \alpha_k (\phi(x_k))^T \phi(x_k) = \sum_{i=1}^{n} \alpha_i k(x_i, x_j)$$

(10)

We can extract the first m (1<m<N-1) nonlinear principal components corresponding to first m non-increasing eigenvalues of (5) using the kernel function without the expensive operation that explicitly projects samples to high dimensional space R^f .

There are some usually used kernel functions as in the following:

Kernel	$k(x_1, x_2)$
Polynomial	$(x_1^T x_2)^d$
Gaussian (radial basis function)	$e^{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}}$
Sigmoid	$\tanh(a x_1^T x_2 + b)$

Table 1. Some kernel functions can be used in Kernel PCA.

3 Kernel Fisher Linear Discriminate

Denoting the within-class S_w^{ϕ} and between-class S_b^{ϕ} , and applying FLD in kernel space, thus we need to find eigenvalues and eigenvectors of

$$\lambda S_w^{\phi} W^{\phi} = S_B^{\phi} W^{\phi} \quad (11)$$

which can be obtained by

$$W_{OPT}^{\phi} = \arg\max_{W^{\phi}} \frac{\left|(W^{\phi})^T S_B^{\phi} W^{\phi}\right|}{\left|(W^{\phi})^T S_w^{\phi} W^{\phi}\right|}$$
$$= \arg\max_{W^{\phi}} \frac{\left|\alpha KZK\alpha\right|}{\left|\alpha KK\alpha\right|}$$
$$= [w_1^{\phi}, w_2^{\phi}, \ldots, w_m^{\phi}]$$

(12)

III. Expeirment and Result analysis

The goal of anomaly identification is to separate the nodes into two classes: those that contain faults (i.e., anomalies), and those that do not. *Sensitivity* and *specificity* are two widely used metrics to measure a binary classification test. *Sensitivity* measures the proportion of actual *positives* which are correctly identified; and *specificity* measures the proportion of *negatives* which are correctly identified. More specifically, they are defined as:

(1) *Sensitivity* is the proportion of correct faulty classifications to the number of actual faults.

Sensitivity = T_P / (T_P + F_N)

(2) *Specificity* is the proportion of correct non-faulty classifications to the number of actual normal nodes.

Specificity = T_N / (F_P + T_N)

Here, T_P, F_P , F_N, and TN denote the number of true positives, false negatives, false positives, and true negatives respectively.

A good method should provide a high value (close to 1.0) for both metrics. A *sensitivity* of 1.0 means that the method recognizes all the faulty nodes, and a *specificity* of 1.0 means this method verifies all the healthy nodes.

We conduct two sets of experiments:

(1) *single-fault tests* where faults of one type are injected into the system;

Fault	ICA-based	PCA-based	KPCA-based	KFDA-based
memory	1.00/0.97	1.00/0.93	1.00/0.98	1.00/0.985
network	1.00/0.93	0.89/0.86	1.00/0.94	1.00/0.97

Accuracy results in case of single-fault tests. There are two numbers in each cell: the first is sensitivity and the second is specificity. A value of 1.00/1.00 indicates a perfect result. And the threshold TH is set to 2.1 in outlier detection.

(2) *multi-fault tests* where multiple types of faults are injected into the system.

Fault	ICA-based	PCA-based	KPCA-based	KFDA-based
CPU&network	1.00/0.94	0.52/0.93	1.00/0.96	1.00/0.97

Accuracy results in case of multi-fault tests. There are two numbers in each cell: the first is sensitivity and the second is specificity. A value of 1.00/1.00 indicates a perfect result. And the threshold TH is set to 2.5 in outlier detection.

Our experimental results show that Kernel Eigenface and Fisherface methods are able to extract nonlinear features and achieve lower error rates. One explanation for some of the superior performance of Kernel Fisherface method over SVM-based method may be attributed to the fact that Kernel Fisherface method uses all training samples to extract the most discriminate (nonlinear) features in the solution.

In terms of execution time, our experiments (with Matlab implementations) show that the ratio of computation loads required by these methods is, on the average, ICA: KFDA: KPCA: PCA = 8.7: 3.3: 3.2: 1.0 (averaged over all the experiments). Obviously, ICA, Kernel PCA and Kernel FDA are all computationally more expensive than PCA. This might be explained that in order to get higher ration of specificity and sensitivity, we need to sacrifice some time to do it. It could be seen as the compromise of efficiency and computation complexity. And also we could easily found that KPCA and KFDA have higher ratio of specificity and sensitivity than ICA, but the computation complexity is lower than ICA, thus KPCA and KFDA are better than ICA.

IV. CONCLUSION AND DISCUSSION

The representations in the classical PCA approach is based on second order statistics of the data set, i.e., covariance matrix, and does not use high order statistical dependencies such as the relationships among three or more pixels. For anomaly detection, much of the important information may be contained in the high order statistical relationships among the pixels. Using the kernel tricks, we extend the classical methods to kernel space where we can extract nonlinear features among three or more pixels. In addition, the ICA method is effective for non-Gaussion distributed data set, but does not work well for Gaussion cases. In this paper, we investigated Kernel Eigenface and Kernel Fisherface methods, and demonstrated that they provide a more effective representation for anomaly detection. Compared to other techniques for nonlinear feature extraction, kernel methods have the advantages that they do not require nonlinear optimization, but only the solution of an eigenvalue problem. Experimental results on two kinds of databases show that Kernel Eigenface and Kernel Fisherface methods achieve lower error rates than the ICA and PCA approaches in anomaly detection.

REFERENCES

[1] Ziming Zheng, Yawei Li and Zhiling Lan. "Anomaly Localization in Large-Scale Clusters."

[2] Thanda Shwe, Win Aye. "A Fault Tolerant Approach in Cluster Computing System." Proceedings of ECTI-CON 2008.

[3] I. S. Jacobs, C. P. Bean and Alexander V. Mirgorodskiy, Naoya Maruyama, Barton P. Miller. "Problem Diagonis in Large-Scale Computing Environments." SC2006 November 2006, Tampa, Florida, USA.

[4] Nicol N. Schraudolph Simon G¨unter S.V. N. Vishwanathan; "Fast Iterative Kernel PCA"; 8(Aug):1893--1918, 2007 IEEE

[5] Alzate, C.; Suykens, J. A. K. ; "Kernel Component Analysis Using an Epsilon-Insensitive Robust Loss Function"; IEEE Transactions on Neural Neworks Volume: 19 Issue: 9, Sept. 2008Pages: 1583-1598

[6] Takiguchi, T.; Ariki, Y.; "Robust Feature Extraction using Kernel PCA"; Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on Volume 1,14-19 May 2006

[7] Martin, S.; "An Approximate Version of Kernel PCA"; Machine Learning and Applications, 2006. ICMLA '06. 5th International Conference on Dec. 2006 Page(s):239 – 244

[8] Dambreville, S.; Rathi, Y.; Tannenbaum, A.; "A Framework for Image Segmentation Using Shape Models and Kernel Space Shape Priors"; Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 30, Issue 8, Aug. 2008 Page(s):1385 – 1399

[9] Su Yan,; Jiu-Fen Zhao,; Jiu-Ling Zhao,; Qing-Zhen Li,; "A method for image classification based on Kernel PCA", Machine Learning and Cybernetics, 2008 International Conference on Volume 2, 12-15 July 2008 Page(s):718 – 722

[10] Z. Ferdousi; A. Maeda; "Unsupervised Outlier Detection in Time Series Data"; Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on 2006 Page(s):x121 - x121

[11] Jiong Zhang; Mohammad Zulkernine; "Anomaly Based Network Intrusion Detection with Unsupervised Outlier Detection"; Communications, 2006 IEEE International Conference on Volume 5, June 2006 Page(s):2388 – 2393

[12] Brugger, D.; Bogdan, M.; Rosenstiel, W.; "Automatic Cluster Detection in Kohonen's SOM"; Neural Networks, IEEE Transactions on Volume 19, Issue 3, March 2008 Page(s):442 – 459

[13] Cuiling Li; Hao Zhang; Jian Wang; Rongyong Zhao; "A New Pattern Recognition Model based on Heuristic SOM Network and Rough Set Theory"; Vehicular Electronics and Safety, 2006. ICVES 2006. IEEE International Conference on 13-15 Dec. 2006 Page(s):45 – 48

[14] Chi Kim Chow; Shiu Yin Yuen; "Signal Self Organizing Map"; Neural Networks, 2007. IJCNN 2007. International Joint Conference on 12-17 Aug. 2007 Page(s):213 - 218

[15] Zhenya Zhang; Hongmei Cheng; Shuguang Zhang; "Approach to SOM based correlation clustering"; Control and Decision Conference, 2008. CCDC 2008. Chinese 2-4 July 2008 Page(s):2485 – 2489

[16] Hujun Yin; "Self-Organizing Map as a Natural Kernel Method"; Neural Networks and Brain, 2005. ICNN&B '05. International Conference on Volume 3, 13-15 Oct. 2005 Page(s):1891 – 1894.

[17] B. Moghaddam. Principal manifolds and bayesian subspaces for visual recognition. In Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 1131-1136, 2005.

[18] A. Durate, D. Rexachs, E. Lugue, "Distributed Scheme for Fault-Tolerant in Large Clusters of Workstations." Parallel Computing: Current & Future Issues of High-End Computing, Proceedings of the International Conference ParCo 2005, NIC Series, Vol. 33, ISBN 3-00-017352-8, pp. 473-480, 2006.

[19] Vo Dinh Minh Nhat, SungYoung Lee; Kernel-based 2DPCA for Face Recognition. IEEE International Symposium on Signal Processing and Information Technology, 2007.

[20] Benítez-Pérez H., García-Nocetti F.; "AN APPROACH FOR FAULT LOCALIZATION BASED UPON UNSUPERVISED NEURAL NETWORKS"; Proceedings of the 2005 IEEE Conference on Control Applications Toronto, Canada, August 28-31, 2005