When Multivariate Forecasting Meets Unsupervised Feature Learning - Towards a Novel Anomaly Detection Framework for Decision Support

Journal:	*International Conference on Information Systems 2012*
Manuscript ID:	ICIS-0481-2012.R1
Track:	16. Knowledge Management and Business Intelligence
Keywords:	Decision Support Systems (DSS), Decision models, Machine learning

Abstract:	Many organizations adopt information technologies to make intelligent decisions during operations. Time-series data plays a crucial role in supporting such decision making processes. Though current studies on time-series based decision making provide reasonably well results, the anomaly detection essence underling most of the scenarios and the plenitude of unlabeled data are largely overlooked and left unexplored. We argue that by using multivariate forecasting and unsupervised feature learning, these two important research gaps could be filled. We carried out two experiments in this study to testify our approach and the results showed that decision support performance was significantly improved. We also proposed a novel framework to integrate the two methods so that our approach may be generalized to a larger problem domain. We discussed the advantages, the limitations and the future work of our study. Both practical and theoretical contributions were also discussed in the paper.

WHEN MULTIVARIATE FORECASTING MEETS UNSUPERVISED FEATURE LEARNING - TOWARDS A NOVEL ANOMALY DETECTION FRAMEWORK FOR DECISION SUPPORT

Research-in-Progress

Jiawei Wang
Department of Information Systems
USTC-CityU Joint Advanced Research Centre
166 Ren'ai Road, Suzhou, China
wangjw2@mail.ustc.edu.cn

Jimmy SJ. Ren
Department of Information Systems
City University of Hong Kong
83 Tat Chee Avenue, Hong Kong
sjren2@student.cityu.edu.hk

Wei Wang
Department of Information Systems
City University of Hong Kong
83 Tat Chee Avenue, Hong Kong
wewang8@student.cityu.edu.hk

Xin Li
Department of Information Systems
City University of Hong Kong
83 Tat Chee Avenue, Hong Kong
Xin.Li@cityu.edu.hk

Qiudan Li
State Key Laboratory of Management and Control for Complex Systems， Institute of Automation, Chinese Academy of Sciences, Beijing, China
qiudan.li@ia.ac.cn

Stephen SY Liao
Department of Information Systems
City University of Hong Kong
83 Tat Chee Avenue, Hong Kong
issliao@cityu.edu.hk

Abstract

Many organizations adopt information technologies to make intelligent decisions during operations. Time-series data plays a crucial role in supporting such decision making processes. Though current studies on time-series based decision making provide reasonably well results, the anomaly detection essence underling most of the scenarios and the plenitude of unlabeled data are largely overlooked and left unexplored. We argue that by using multivariate forecasting and unsupervised feature learning, these two important research gaps could be filled. We carried out two experiments in this study to testify our approach and the results showed that decision support performance was significantly improved. We also proposed a novel framework to integrate the two methods so that our approach may be generalized to a larger problem domain. We discussed the advantages, the limitations and the future work of our study. Both practical and theoretical contributions were also discussed in the paper.

Keywords: Decision support systems (DSS), decision models, machine learning

Introduction

Nowadays information technology endows many organizations and governmental agencies the possibility of making intelligent decisions during operations. Among the types of data which support such decision making process, time-series data is widely used due to its availability and accessibility. Time-series data reflects the dynamics of operations or status over time and contains rich information which can be unveiled to support various aspects in the decision making process. Many examples of using time-series data in decision support can be found in the extant information systems (IS) studies including identifying traffic incidents for transport authority via transportation IS (Yuan et al. 2003), labeling customers into different categories for supporting company policy making by marketing IS (Lee et al. 2012), differentiating patients in supporting healthcare decision making by medical IS (Yeh et al. 2011), etc. As there is no universal term to summarize these studies, we collectively name this research field time-series data enabled (TDE) decision support.

The methodology underlying most of the existing studies in this field is to formalize the decision support scenario into a standard classification problem (e.g. to classify whether a firm has a high risk of bankruptcy) and process the data by using a classifier (Athavale et al. 2009). Classification models or rule based classifiers trained by historical data may be used in differentiating states or events for decision support. While this approach works reasonably well in many applications, we argue that a common characteristic of these decision support scenarios is overlooked by most of the previous studies. That is in the majority of context which belongs to TDE decision support, the number of samples in one class is usually much smaller than that in the rest. We can easily discern this phenomenon in the examples we just mentioned. Take labeling customers of telecommunication companies as an example, samples of customers who cancel the contract are always scarce. The same situation is almost always true in other contexts such as healthcare decision support and transportation decision support, etc. Therefore, we believe most of the scenarios in TDE decision support can actually be formalized as anomaly detection problems rather than just standard classification problems.

In addition, in all the previous studies only labeled data was used in the training of the classification model. While this is a widely used approach, we can’t ignore the fact that labeling much time-series data is often very expensive and sometime impractical. Furthermore, sufficient unlabeled data is always available and easy to obtain but unfortunately left unused. For instance, in fraud transaction classification we might not have many labeled fraud transaction samples in the data, but as long as the financial IS is up and running, the amount of transaction data (without knowing whether it’s a fraud or not) is always growing. Therefore, the ability of taking advantage of the huge amount of unlabeled data in the TDE decision support, if such approach is proven successful, is an alternative way to improve performance without introducing any extra cost and risk to the existing system simply because much unlabeled data already exists in the system.

The subtlety of these two new perspectives is twofold. First of all, if anomaly detection as well as the ability to utilize unlabeled data is a better way to approach TDE decision support, the potential of improving the state-of-the-art performance is worth exploring. Secondly, as a result of the richness of different scenarios in the TDE decision support context, a novel anomaly detection based framework for this particular context may be established and provide theoretical insights for real life applications. In this paper, we propose to solve the TDE decision support problems from these two perspectives, namely anomaly detection and the utilization of the plentiful unlabeled data. A novel anomaly detection framework will also be established to integrate these two perspectives for decision support.

The structure of the paper is as the following. The second section reviews the literature in decision support and identifies the research gaps. The third section describes the proposed framework and the research methodologies. The fourth section demonstrates the effectiveness of the proposed methodology and presents the preliminarily experiment results. We summarize the limitations of this study, discuss and propose the future work in the last section.

Literature Review

In the previous research on decision making in IS, many studies can be categorized into TDE decision support. Stock prices were analyzed for identifying potential business failure by means of kernel based machine learning approach in three sectors for assisting investors in making a right strategy (Athavale et al. 2009). Incidents were detected automatically via support vector machine (SVM) to enable a quick incident response and traffic control decision making in (Yuan et al. 2003). Time-series based patient biochemical measures were utilized to classify hemodialysis patients into the category of requiring hospitalization according to rules extracted from historical time-series biochemical data for decision making in medical resource arrangement (Yeh et al. 2011). Lee et al. (2012) demonstrated the way of finding customers intending to cancel the contract by using k-nearest neighbor algorithm for telecommunication companies in making retention policy.

Generally speaking, though most of the studies in TDE decision support solve this problem from a standard classification perspective (either model based classification or rule based classification) the underlying scenarios of these studies significantly overlap with time-series based anomaly detection. More concretely, in the retention policy making problem stated in (Lee et al. 2012), the number of customers tentatively cancel the contract is relatively rare compared to the staying customers. Incidents happening on the road segment are defined to be non-recurrent event so that the probability of incident occurring is low (Dia et al. 2011). Abnormal patterns disclosed in biochemical measures of Hemodialysis patients requiring hospitalization are also rare (Yeh et al. 2011). The aforementioned examples target to carry out decision support for the events that are unusual, unpredictable, and relatively rare. According to (Chandola et al. 2009), anomalies are patterns in data that are not consistent with the notion of normal behavior and they are sufficiently far away from normal conditions, and they are rare. Therefore, we argue that TDE decision support can be conceptualized as anomaly detection for decision support.

From the existing TDE decision making literature, it's also clear that though the amount of unlabeled data is almost always plenty and ready to be used, the utilization of unlabeled data is largely overlooked. In the incident detection case (Yuan et al. 2003), unlabeled traffic data is constantly generated by underground sensors minute by minute and stored in the data center. However, only labeled data which actually occupies a tiny part of the whole dataset available is used in the training and testing phases. Same situation happens in the marketing decision support scenario (Lee et al. 2012), the health care decision support scenario (Yeh et al. 2011) and many others. If there is an effective way to recognize and embrace the usefulness and importance of this yet unexplored resource, it would be much preferable because we would be able to add values without introducing extra effort in data collection.

The two new perspectives above well illustrate the research gaps in the existing TDE decision support literature, the overlook of the underlying anomaly detection essence and the ignorance of unlabeled data. In order to seek the possible solutions to realize the two perspectives just identified and supported by the literature, we first turn to the literature of time-series based anomaly detection to investigate the methodology which may be used in TDE decision support. The most systematic methodology we found is the combination of a forecasting component and a detection/monitoring component (Lotze 2009; Lotze et al. 2009; Yang 2011). The function of the forecasting component is to forecast the time-series with a normal condition time-series model and the detection/monitoring component focuses on exploring the discrepancy between the forecasted condition and actual condition for anomaly detection. Additionally, Lotze et al. (2009) indicated that utilizing related indicators leads to an improvement in forecasting and a corresponding improvement in detection. Due to the solid theoretical foundation and the promising performance, multivariate forecasting is usually adopted in many complex situations. Among the techniques for the detection component which are summarized in (Chandola et al. 2009), we found that classification based techniques conform with the techniques used in most applications of TDE decision support. Based on the above discussion, we believe multivariate forecasting tied with a classification detection component is a promising unified approach to conduct the anomaly detection for TDE decision support.

Then we would like to answer the question namely, what method is the most appropriate one to take advantage of unlabeled data in the TDE decision support process. There is an emerging research field in machine learning collectively named deep learning and unsupervised feature learning provides a promising solution to this question (Bengio 2009). When generating features for classifiers, the common approach is either to use the raw data or handcraft the features which are appropriate for the particular task. Unsupervised feature learning takes a different approach. It uses unsupervised learning methods to seek internal representations of the raw data and build a higher level representation automatically from it (Adam Coates et al. 2011; Bengio 2009). Since the whole procedure is purely unsupervised, no labeled data is required in building the feature mapping function. Studies showed that unsupervised feature learning was able to exceed state-of-the-art performance in many classification tasks including anomaly detection tasks (Bengio 2009; Lee et al. 2009). Though not many studies were devoted to explore the power of unsupervised feature learning in processing time-series data, the underpinning concept of this method well fits our context and therefore worth an extensive exploration.

For the purpose of generalizing our new perspectives and providing solid external validity of this study, a unified framework is proposed to illustrate the generalized approach for TDE decision support problem. In the framework we connect the forecasting component, detection component, and the unsupervised feature learning component. To our knowledge, there is no previous study attempting to carry out this generalization and therefore our framework is novel.

Methodology

To bridge the aforementioned research gap, our research aims to design a generalized approach for supporting decision making by detecting anomalies from multiple time-series indicators of which lots of them are unlabeled. Therefore, we propose a novel anomaly detection framework for decision support which is shown in figure 1. The multivariate forecasting and unsupervised feature components are the core parts of the framework. By adopting multivariate forecasting component for normal condition, we hope this forecasting component is able to enhance the anomaly detection performance since the anomaly in nature deviates sufficient from the forecasted normal condition. We also introduce the unsupervised feature learning for generating higher level features from the real-time condition and its normal forecasting for anomaly detection.

Novel Anomaly Detection Framework for Decision Support

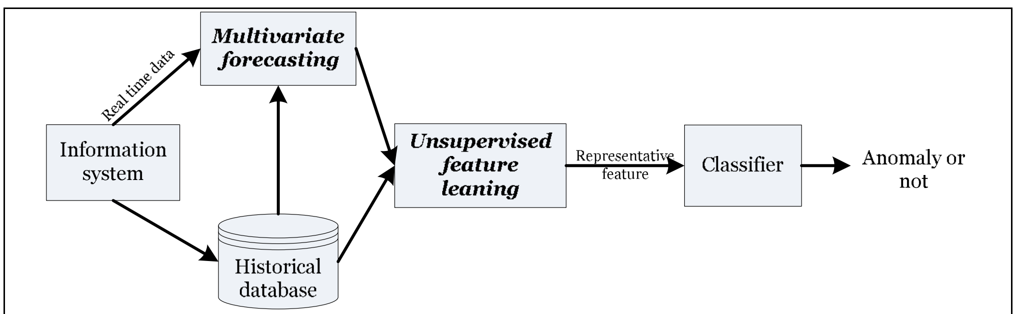

Figure 1. Anomaly Detection Framework for Decision support

There are five components of the proposed framework. Information system is responsible for collecting and storing the multiple time-series data. Of the historical database, data under normal condition is used to train the multivariate forecasting model. Data points before the real time data use the saved multivariate model for forecasting conditions under the normal assumption. Historical data without labels, together with their normal expected normal condition (the output of multivariate forecasting module), are inputted into the unsupervised learning component for generating the feature mapping function. Then the feature mapping function is then used to map the raw feature comprising real time data and the corresponding forecasting result into a more effective representative feature. Finally, the higher level feature is fed into the anomaly classifier for anomaly detection.

Among the components in the proposed framework, the classifier module has been studied relatively extensively. Thus this component is relatively mature. Information system and the historical database can also be implemented effectively by mature technologies. However, the identified research gaps are to be fulfilled by the multivariate forecasting module, the unsupervised feature learning module and the combination of these two. Multivariate forecasting is proven to be useful in some anomaly detection scenarios (Lotze et al. 2009). We introduce it in our framework to explore its effectiveness in TDE decision support and to bridge the first research gap we aforementioned. As to the second gap, the ignorance of plentiful unlabeled data in TDE decision support scenarios, unsupervised feature learning is adopted in the framework to utilize the unlabeled data and to generate more effective higher level features.

Multivariate Forecasting

We formulate the normal condition forecasting as a multivariate forecasting problem. Effective multivariate time-series (MTS) data modeling is critical for many decision making activities (Trapero et al. 2012). Therefore much research has been conducted on analyzing MTS data from both statistical and artificial intelligence perspectives. Approaches for statistical MTS modeling include the Vector Auto-Regressive (VAR) process, the Vector Auto-Regressive Moving Average (VARMA) process, and some other non-linear and Bayesian approaches (X Liu et al. 1999).

VAR model is developed for multivariate time-series forecasting and this model has been proven to be especially useful for capturing the dynamic behavior of time-series data in economic, financial (Ang and Piazzesi 2003) and traffic (Chandra and Al-Deek 2009) area. Therefore, we adopt this model to forecast the normal condition or state in this study. Assume we have a vector of time-series variables which is denoted as $X = (x_{1,} x_2, ..., x_i, ..., x_m)$ for $i = 1, ..., m$. The estimation of vector X by VAR model is $Y = (y_{1,} y_2, ..., y_m)$. The general VAR formula is as the following.

$$Y_t = A_1 X_t + A_2 X_{t-1} + ... + A_j X_{t-j} + ... + A_p X_{t-p} + u_t \quad (1)$$

Where A_j are (m*m) coefficients matrices for $j = 1, ..., p$ and u_t is a m-dimensional process with $E(u_t) = 0$ and time invariant positive definite covariance matrix $E(u_t u_t^T) = \sum_u$ (white noise).

In order to obtain Y_t under the normal condition, the model should be built based on the historical data of normal condition. The number of time lags for VAR model is selected based on Akaike's information criterion (AIC) and the coefficients of the model are tuned by using ordinary least squares (OLS).

Unsupervised Feature Learning

Unsupervised feature learning discovers internal structures within the unlabeled raw data and forms a feature mapping function. Then this mapping function is applied to raw features to form higher level features for anomaly detection tasks. The unsupervised feature learning algorithm is briefly described in the following. Step 1, to create an array with a number of $z+1$ dimensional unlabeled raw data vectors from the specified training data ($z = 12$ in this study). Step 2, to extract random patches from unlabeled raw data vectors. Step 3, to use an unsupervised feature learning algorithm to learn a feature mapping function from the patches. Step 4, to prepare $z+1$ dimensional raw features from the training set. Step 5, to apply the learnt feature mapping function to sub-patches within each $z+1$ dimensional raw feature representation of every labeled detection moment. Then to generate the higher level features for labeled data using pooling (Boureau et al. 2010). We repeat step 1 to step 5 when necessary, e.g. do it for different indicators respectively. In step 6, we then enhance the raw features with the newly generated higher level features and use the enhanced features to train a classifier to classify anomalies.

In step 3, we apply K-means clustering to learn K centroids c^k from the input data. We purposefully used K-means algorithm in this study. K-means is not only an effective unsupervised feature learning algorithm (Lee et al. 2009), it is also straightforward to implement. Moreover, unlike other methods, K-means has only one parameter (number of centroids) to be fine-tuned in the feature learning stage. This fact significantly saves the training time. Therefore K-means algorithm is the ideal choice to work with in this study. Given the learnt centroids, we construct the following feature mapping function:

$$f_k(x^{,}) = \max\{0, \mu(\tau) - \tau_k\} \quad (2)$$

where $\tau_k = \| x^{,} - c^k \|_2$, $\mu(\tau)$ is the average of all the values in τ , the input of the feature mapping function $x^{,}$ is the raw feature for a detection moment to be prepared in step 4. The output of the feature mapping function is a K dimensional higher level feature vector in which the kth element is zero if the distance between $x^{,}$ and c^k is above average. This feature mapping function will generate higher level features with some degree of sparseness.

Experiments and Preliminary Results

In order to demonstrate the validity of the proposed framework, three steps are required. First, the effectiveness of the two crucial components: multivariate forecasting and unsupervised feature learning need to be testified separately. Second, the effectiveness of the whole framework should also be testified. Third, the generalisability of the framework namely, the ability to apply the framework in various applications in TDE decision support should also be verified.

In this stage of the research, we focus on step one. For the multivariate forecasting component and the unsupervised feature learning component, we design two experiments respectively to examine their potential in enhancing the decision support performance by combining this component with a classifier. Traffic incident dataset was used to implement the two experiments mentioned above. There are three reasons that we choose traffic incident data in this experiment. Firstly, incident is a typical anomaly in routine transportation and it is a critical problem in modern transportation systems and operations. Incidents cause congestion on the road and reduce the efficiency of transportation systems. Essentially, detecting incidents is similar to the process of detecting credit card fraud and identifying telecommunication customers intending to cancel the contract from regular business transactions. Secondly, there are a large number of multivariate time-series data of different frequencies for road segments and the labeled data is relatively expensive to get. Thirdly, like many scenarios in business, transportation agencies also need effective decision support when making decisions on choosing appropriate incident response strategies when there is an incident occurs on the road segment. In sum, traffic incident detection is a classical time-series based anomaly detection problem for supporting decision making, and hence the incident data is suitable in our context.

For the first experiment, forty-nine days traffic data (including volume, occupancy, and speed) was collected from I-880 freeway (Petty 1996) in California. Traffic data was averaged over lanes in the temporal aggregation of 1 minute. There are 45 lane-blocking incidents in the first experiment. For the second experiment, real incident data were collected in two sites namely I-405 northbound freeway and SR-22 eastbound freeway in the Orange County, California. The dataset contains 181 lane-blocking incidents in the entire year of 1998 and includes both traffic volume and occupancy data averaged within the interval of 30 seconds.

For evaluating the proposed idea in incident detection problem, three kinds of widely accepted metrics (Singliar et al. 2010) are selected in this study. The first metric is the detection rate (DR) which is the ratio of number of detected incident cases to the total number of incidents in the dataset. The second metric is the false alarm rate (FAR), which is defined as the ratio of falsely identified incident alarm cases to the total number of instances. The last metric is the mean time to detect (MTTD), which is the total time for actually detecting incidents divided by the number of detected incidents.

In the incident detection problem, one incident detection algorithm that can give a high DR, low FAR, together with a short MTTD outperforms other approaches. Performance curves of DR vs. FAR and MTTD vs. FAR are usually employed to compare the performances of different incident detection algorithms.

In the first experiment, VAR model for forecasting normal traffic variables is associated with SVM classifier to detect incidents. We name this approach as VAR-SVM in our study. In the second experiment, an unsupervised feature learning component based on K-means clustering is selected for extracting high-level representative features. The new features are then fed into a SVM to detect incidents.

Experiment Ⅰ

For each detector station of the road segment, VAR model for expected normal traffic estimation is built based on historical traffic patterns under normal condition. The model incorporates traffic (volume, occupancy, and speed) from its downstream and upstream detector stations. Package 'vars' in R language (Pfaff 2008) was used to select the number of time lags for VAR models by optimizing AIC. For each time point, data from upstream or downstream include real-time traffic, expected normal traffic, and the difference between these two. In order to classify a road segment status at time *t* as an incident or normal condition, according to (Yuan et al. 2003), upstream data from *t*-4 to *t* and downstream data from *t*-2 to *t* were fed into SVM. We used 5-fold cross-validation to tune parameters for SVM.

Results for experiment Ⅰ are presented in Fig. 2. From the performance curve of DR vs. FAR, it is clear that VAR-SVM has a higher DR than SVM baseline while FAR is less than 0.8%. Moreover, VAR-SVM detects an incident more quickly than SVM when FAR is in the range of [0, 0.8%]. Though the two approaches detect the same number of incidents when FAR is more than 0.8%, it is evident that VAR-SVM detect an incident in less number of minutes than SVM does which is shown in the performance curve of MTTD vs. FAR. In sum, VAR-SVM outperforms SVM baseline in DR and MTTD. In light of the better performance of VAR-SVM, we believe that it is of great potential to enhance the performance of supporting decision making from the perspective of detecting anomalies from time-series data.

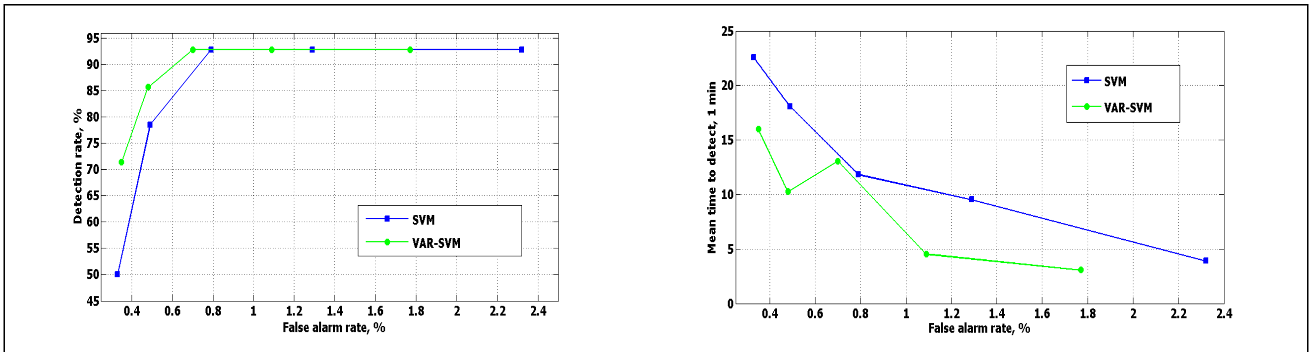

Figure 2. Preliminary result of multivariate forecasting

Experiment Ⅱ

Similar setup as in experiment Ⅰ was adopted in experiment Ⅱ. We carried out the unsupervised feature learning process described in the methodology section. Due to the randomness of K-means algorithm, we repeated the experiment for 50 times. The final results were obtained by averaging the 50 results. The number of centroids we used was 75 for volume, 15 for occupancy; patch size was 11 for volume, 6 for occupancy; number of patches sampled was 20000. Therefore, the dimension of the learnt higher level features was 180 (75*2+15*2). Due to the randomness of patch sampling and K-means, SVM parameter selected by cross validation was usually different among 50 times. The following diagram shows the preliminary performance of incident detection by using unsupervised feature learning.

Multiple experiments with multiple choices of raw features and new features were conducted. Due to the limitation of the space, we only list one of the typical results. A more comprehensive description of this part of the experiment can be found in our previous work (Ren et al. 2012). Several observations can be summarized from the results. First, by using the enhanced features, significantly better (higher) DR was obtained when FAR is ranged 0.1% to 0.7%. Second, strictly better (lower) MTTD was also obtained for a large range of FAR by using the enhanced features. Therefore, we can conclude that the enhanced feature leads to significantly better overall incident detection performance. As the only factor we manipulate in this experiment is whether to use enhanced features or raw features in the anomaly detection task, we are also confident to say that the utilization of unlabeled data is successful by using unsupervised feature learning.

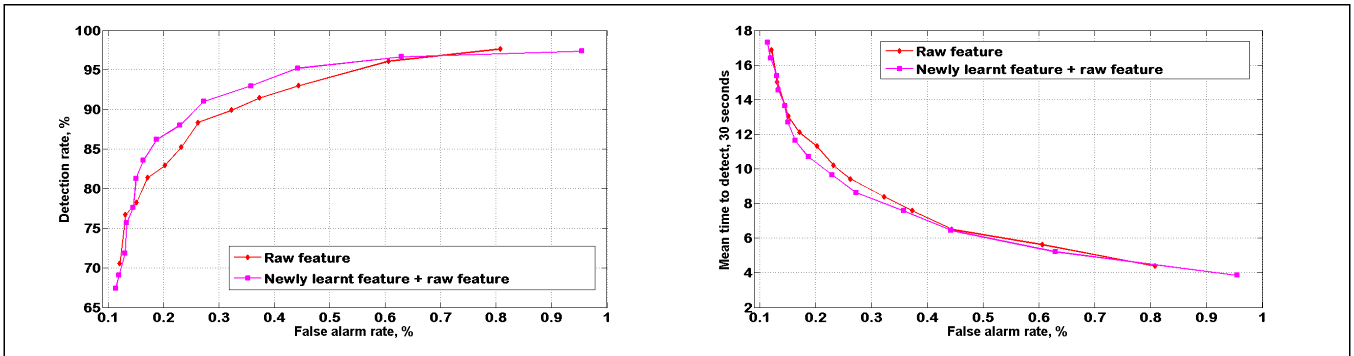

Figure 3. Preliminary result of unsupervised feature learning

Conclusion and Future Research

Discussion

In this study, we propose a novel anomaly detection framework for supporting decision making in IS. The idea that formalizes decision support as a problem of time-series based anomaly detection provides us a novel way of supporting decision making. In order to solve the anomaly detection problem with limited amount of labeled data, we introduced the multivariate forecasting and unsupervised feature learning components into the proposed framework.

We conducted two experiments to test the effectiveness of two critical components of the proposed framework on traffic incident datasets. The first experiment's result showed that the performance of detecting incidents was enhanced with the inclusion of the multivariate forecasting capability. By analyzing the outcome of the second experiment, we found that the common concern about the scarcity of labeled time-series data in anomaly detection can be dealt with by utilizing an unsupervised feature learning technique. Based on the promising performance we gained from the two experiments, we believe that the proposed framework stands a good chance of supporting decision making.

However, there are limitations in the study. In this phase of the study, the two components were testified separately so that the effectiveness of the whole framework remains unknown. Secondly, the criteria of selecting particular combination of multivariate forecasting technique and classifier are not determined. Thirdly, this framework is put forward for using time-series data for supporting decision making while there are considerable amount of data of other types in real life applications. Other information such as expert experiences is worthy of being considered in the decision making process.

Contributions

There are two contributions in this study. In the practical side, by identifying the underlying anomaly detection nature and the overlook of the unlabeled data in TDE decision support, we utilized multivariate forecasting and unsupervised feature learning to refine the existing work. We were able to successfully improve the performance of TDE decision support from two complementary perspectives namely, improving the existing classification technique and enlarging the existing training resource. On the other side, the richness of the TDE decision support scenarios and the lack of a systematic method which emphasizes their anomaly detection essence call for a comprehensive framework to integrate different components in the decision support process. We response to this demand by proposing a framework comprising components deal with various important aspects within the decision support process. The proposed framework shall provide the theoretical guidance based on which other researchers in the field may build their own decision support systems. It may also contribute to form the cornerstone of summarizing the systematic viewpoint in dealing with TDE decision support problems.

Future Work

Motivated by the limitations, certain number of future research directions should be carried out. The performance of the whole framework should be testified by more datasets. Though VAR model and K-mean algorithm were selected in this study, the future study should testify other possible methods for multivariate forecasting and unsupervised feature learning. Besides, the relationships between the components are also required to be explored. By doing so, the revealed relationship could shed light on the way of selecting appropriate implementation techniques for the components in the framework.

The above mentioned design and finished work are done on the incident dataset for decision making in transportation management, it would be necessary to apply the framework to other domains such as networking monitoring, customer retention policy making, stock investment, bank transaction management to testify the generalisability of the proposed framework.

Acknowledgements

Research reported in this paper was supported by Innovation and Technology Fund of Hong Kong SAR government (ITP/039/09AP: A Technological Platform for Traffic Information Collection and Integration), CityU StUp7200170, and SRG 7002625. We thank anomalous reviewers and the editors for providing us valuable comments and suggestions.

References

Adam Coates, Honglak Lee, and Y.Ng., A. 2011. " An Analysis of Single-Layer Networks in Unsupervised Feature Learning. International Conference on Artificial Intelligence and Statistics," *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*.

Athavale, Y., Hosseinizadeh, P., Krishnan, S., and Guergachi, A. 2009. "Identifying the potential for failure of businesses in the technology, pharmaceutical and banking sectors using kernel-based machine learning methods," *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pp. 1073-1077.

Bengio, Y. 2009. "Learning Deep Architectures for AI." *Foundations and Trends in Machine Learning* (2:1), pp 1-127.

Chandola, V., Banerjee, A., and Kumar, V. 2009. "Anomaly detection: A survey," *ACM Comput. Surv.* (41:3), pp 1-58.

Dia, H., and Thomas, K. Jan 2011. "Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data," *Information Fusion* (12:1), pp 20-27.

Lee, H., Largman, Y., Pham, P., and Y.Ng, A. 2009. "Unsupervised feature learning for audio classification using convolutional deep belief networks," *Neural Information Processing Systems (NIPS)*.

Lee, Y.-H., Wei, C.-P., Cheng, T.-H., and Yang, C.-T. 2012. "Nearest-neighbor-based approach to time-series classification," *Decision Support Systems* (53:1), pp 207-217.

Lotze, T. H. 2009. "Anomaly detection in time series: theoretical and parctical improvements for disease outbreak detection," *Ph.D. dissertation*, *University of Maryland*.

Lotze, T. H., and Shmueli, G. 2009. "How does improved forecasting benefit detection? An application to biosurveillance," *International Journal of Forecasting* (25:3), pp 467-483.

Petty, K., Noeimi,H.,Sanwal, K., Rydzewski, D., Skabardois, A., Varaiya, P., Al-Deek, H. 1996. "The freeway service patrol evaluation project: database support programs and accessibility" *Tranportation Rsearch* (4C:3), pp 71-86.

Pfaff, B. 2008. "VAR, SVAR, and SVEC Models: Implementation Within R Package vars," *Journal of Sattistical Software* (27:4).

Ren, J, Wang W, Wang JW and Liao Stephen. 2012. "An Unsupervised Feature Learning Approach to Improve Automatic Incident Detection," in *Proceedings of the 15th IEEE International Conference on Intelligent Transportation Systems (ITSC 2012).*

Singliar, T., and Hauskrecht, M. Jun 2010. "Learning to detect incidents from noisily labeled data," *Machine Learning* (79:3), pp 335-354.

Trapero, J. R., Kourentzes, N., and Fildes, R. 2012. "Impact of information exchange on supplier forecasting performance," *Omega* (40:6), pp 738-747.

X Liu, S Swift, A Tucker, G Cheng, and Loizou., G. 1999. "Modelling Multivariate Time Series," *In Proc. of fourth workshop on Intelligent Data Analysis in Medicine and Pharmacology*, pp 201-214.

Y-Lan Boureau , J. P., Yann Lecun 2010. "A Theoretical Analysis of Feature Pooling in Visual Recognition". International Conference on Machine Learning," *27th International Conference on Machine Learning, Haifa, Israel*.

Yang, Y. 2011. "A novel network traffic anomaly detection model based on superstatistics theory," *Journal of Networks* (6:2), pp 311-318.

Yeh, J.-Y., Wu, T.-H., and Tsao, C.-W. 2011. "Using data mining techniques to predict hospitalization of hemodialysis patients," *Decision Support Systems* (50:2), pp 439-448.

Yuan, F., and Cheu, R. L. Jun-Aug 2003. "Incident detection using support vector machines," *Transportation Research Part C-Emerging Technologies* (11:3-4), pp 309-328.