Towards an Adaptive Forecasting of Earthquake Time Series from Decomposable and Salient Characteristics

Simon Fong, Zhou Nannan
Department of Computer and Information Science
University of Macau
Macau SAR
ccfong@umac.mo, ma96578@umac.mo

Abstract—**Earthquake forecasting is known to be a challenging research program for which no single prediction method can claim to be the best. At large, earthquake data when viewed as a time series over a long time, exhibits a complex pattern that is composed of a mix of statistical features. A single prediction algorithm often does not yield an optimal forecast by analyzing over a long series that is composed of a large variety of features. A new analytic framework is proposed that allows these mixed features from the time series to be automatically extracted by a computer program, and fed into a decision tree classifier for choosing an appropriate method for the current forecasting task. The motivation behind this concept is to let the data decide which prediction algorithm should be adopted, adaptively across different periods of the time series. The contribution of this paper is twofold: (1) a framework of automatic forecasting which is very suitable for real-time earthquake monitor is proposed, and (2) an investigation on how different features of the data series are coupled with different prediction algorithms for the best possible accuracy.**

Keywords-Earthquake prediction; time series forecasting; automatic and adaptive forecasting; ARIMA; Holt-Winter's.

I. INTRODUCTION

In literature a large collection of time series forecasting algorithms have been studied and they have been applied for different domains such as finance [1], safety of power system [2], cargo volume [3], and traffic [4], etc. These algorithms show pros and cons in various situations, and mostly they were used for manual analysis where human experts were involved. Forecasting by manual analysis implied a full set of historical data or some intuitively selected length of the data series was used to be experimented by several ad-hoc forecasting algorithms, often trial-and-error or brute-force for the best candidate. This process is not only manual but usually would have to be carried out in the backend that falls short of real-time prediction capability.

Automatic forecasting is more desirable than backend analysis especially for critical scenarios, for instance, real-time monitors for earthquake prediction [5, 6], where fresh inputs from sensors are streamed in continuously and the responsive prediction must be made instantly if a fore-warning were to be made. This kind of real-time prediction scenario would often be handled by a computer program that is capable for processing a large amount of data, extracting characteristics from them and picking a suitable forecasting algorithm plus the required parameters, automatically.

Furthermore, earthquake data are multi-dimensional in nature, so are the prediction tasks. Earthquake data are spatial-temporal events; samples are measures of magnitudes, frequencies, depth, and they are collected from different places at different times. Aggregating up the measurements with other seismological and climate factors and time scales will multiply the dimensions, therefore demands for automatic forecasting for the sake of both speedy and accurate prediction results.

One major obstacle for achieving automatic forecasting is the choice of the appropriate time series forecasting algorithm and the associated parameters for the best results. In the mathematics domain, we are not short of theoretical models for forecasting; rather practically, how each of these different models could be dynamically chosen in run-time based on the characteristics of the latest section of the data series that is being analyzed – this forms the motivation of the research as in this paper.

On the other hand, earthquake data when formatted into a time series, it is found to possess a set of complex features which is often mixed into some composite along a very long range of time (>30 years), with no simple distinct characteristics. Consequently this implies various prediction techniques that have their strengths for different characteristics of patterns should be applied in turn adaptively for each segment of the time series, instead of fixing on a single algorithm that may deem fit at the start for the whole of the historical time series. The rationale of this adaptive prediction model is based on the observations that mixed features reside in earthquake patterns, these features are ever changing over time, and local trends are more relevant than global trends in earthquake. The second observation is still quite debatable because nobody can be certain that global trends have absolutely no effect on us given that the archival dataset only has a history of merely three decades. Local trends however are known to have temporal effect over a period of time and relevant to future events. For example, seismic activities are related in time and over some places, such as fore-shocks and after-shocks.

In this paper, we propose an adaptive forecasting framework for predicting the frequency of earthquakes in future time to come. The advantage of our framework is to facilitate automatic real-time forecasting. Section 2 reviews related works on automatic forecasting. Our proposed framework is described in details in Section 3. Section 4 shows our preliminary experiments. Section 5 concludes.

 ISBN: 978-1-61208-158-8

II. RELATED WORKS

In the literature, tons of forecasting algorithms and methodologies have been studied on univariate time series data from statistical perspectives. The methods are quite mature in terms of accuracy and robustness. Their applications are applied in a wide range of domains for solving real life problems. Automatic forecasting is a relatively unexplored area in software engineering that is comprised of different techniques, in addition to of course existing forecasting algorithms. The challenges of automatic forecasting basically can be divided into three conceptual levels: 1. Model selection, 2. Parameters selection for the chosen model, and 3. Data selection – how long the input data should be used in order to yield a result that has the highest accuracy? Given certain input data, how far ahead in future time intervals the forecast can sustain its accuracy?

With these three levels of complications in automatic forecasting, this section reviews some related works pertaining to practical challenges at each level.

A. *Model Selection*

Most researchers advocated all the popular forecasting models should be first trialed and the one that produced the lowest error is selected in an automatic forecasting approach. The merit of trying out all the models is apparently a way to guarantee the best prediction output, which is as good as using brute-force approach. The drawback of this method is the long computation time required that may be applicable to manual analysis where highest accuracy needs to be assured and an analyst can take a long time to find the perfect model. For real-time forecasting where a very short time constraint is imposed, some heuristic search for the optimal model is needed.

In the 70's, an epitome of using all the forecasting techniques in the hope of finding a technique that is best suited to a certain scenario is called M-Competition [7, 8] by Makridakis et al. It was concluded from M-Competition that statistically sophisticated methods might not necessarily outperform simple forecasting ones. In year 2000 the same authors focused on using only a limited set of exponential smoothing models, and they demonstrated that automatic forecasting can made to be particularly good at short term forecasts [9]; this methodology is extended to seasonal short-term series where the result beats all other methods.

Recently, Hyndman and Khandakar [10] improved this exhaustive M-Competition approach for finding the suitable method in automatic forecasting. They used state space models that underlay exponential smoothing methods for shortening the search process. And the authors also proposed a step-wise algorithm for forecasting with Autoregressive Integrated Moving Average (ARIMA) models.

Lemke and Gabrys in their work [11] suggested identifying an extensive pool of computable features from the time series for choosing a forecasting method. The judgmental feature selection used in [11] was a simple decision tree that decides which forecasting method should be used. Based on this concept, the author in this paper extended the decision tree into one that is for stream mining, treating the input time series data as a continuous stream; thus instead of preprocessing the full set of time series data for constructing (training) up the decision tree, our model progressively builds the decision tree ground up and updates the decision tree dynamically as each new segment of time series arrives. The advantage of using decision tree of stream mining is that the forecasting model can better adapt to different parts of the time series as supposedly the time series should never end but amounts to infinity, and the features of the series ever keep changing in values at different times. The adaptive model is hence suitable for real-time forecasting applications where forecast is made on the ever updating data stream.

B. *Parameter Selection*

Many time-series analysis techniques like regression analysis and exponential smoothing techniques require a set of parameters or constants to be initially chosen by users. The forecast accuracy largely attributes to the choice of these parameters at all times. Basically, there are two schools of researchers on two practices that deal with parameter selection. One common practice [12] is to utilize Solver (an optimization plug-in for MS-Excel) to simultaneously optimize the parameters and smoothing constants in spreadsheet prior to forecasting time series data. Rasmussen in his work [13] optimized using Solver initial or starting parameters as well. This work showed improved fits when the initial parameters and the smoothing constants are optimized together. The other approach [14] is to treat the update equation and the selection of parameters in the forecasting equations as a recursive updating function. The smoothing parameters are optimally chosen by a simple grid search according to some criterion. The solution may fall on local optima as the whole set of data are not globally considered.

In our proposed model, we combine the aforementioned approaches; because of the rolling nature of our forecasting process, the last optimized parameters from the previous segment of the time series would serve as the optimal initial parameters for the current segment by the concept of [13]. The parameters of the forecasting equations for each section of the time series are updated iteratively according to [14].

C. *Data Selection*

The selection of time scales as well as the length of time series has been studied widely in the literature. Generally forecast combinations that are based on segments of the series are said to be superior to their constituent forecasts. The rationale behind is that the conditions that vary along the time series, consists of trend and other statistical elements changes, hence the forecasting parameters drift. Previous works [15, 16, 17] show that it is feasible to do forecasting over different terms of the time series. Sets of forecasts from both short and long terms are merged to outperform either one. Cointegration is the core method adopted in piecing up the forecasts. This underlying concept of forecasting on individual segments instead of a full length of the time series drives the need of an adaptive forecasting model which continuously chooses the best suitable forecasting techniques and parameters on the fly over a train of time series segments.

III. FRAMEWORK FOR ADAPTIVE FORECASTING

A. *Methodology*

Adaptive forecasting is about a continuous forecasting process, which takes the time series input as an ever changing data stream, and automatically chooses a suitable forecasting method based on the characteristics of the recent data. The selection of the forecasting method is adaptive to the time series whose characteristics may vary across different observation time periods.

A classical forecast development process by Armstrong [16] was designed for analyst who makes once-off forecasts from a full length of time series. Our adaptive forecast however is derived from segments of time series of variable length as predefined by the user. The adaptive forecasting process can be implemented as a computer software program; hence human intervention can be relieved making it suitable for automatic and real-time prediction when the source data feed is properly setup.

The methodology of adaptive forecast process which is extended from [18] is shown in Figure 1.

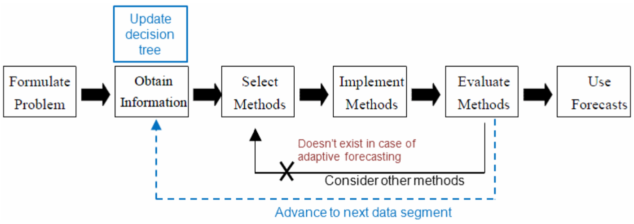


Figure 1. Adaptive forecasting methodology.

The main differences between Armstrong's process and the adaptive process are the use of a decision tree classifier in picking the suitable forecasting method and the time series is inputted to the classifier progressively in a stream of segments. In the original Armstrong's process, the time series is initially studied by identifying the potential explanatory variables. Yaffee [19] rendered the series stationary in mean, variance, and autocovariance; then next step was to manually select a set of forecasting methods from a pool of all available methods, probably by domain knowledge; tried them out, evaluated their performance. The loop repeated by trying the remaining methods if the earlier methods were not performing satisfactorily. In our methodology this loop is eliminated because the manual selection is replaced by a decision tree that instantly finds the appropriate method with optimized parameters given the statistics of the series segment.

The operation of the adaptive forecasting is iterating as the time series rolls along in segments into the classifier. Therefore fresh forecasts are obtained incrementally as time elapses based on the latest data segment. According to the study in [20], it is known that global trend models may not offer the best results. Some researchers advocate that forecasting models that tap on local trends can generally provide far better descriptions of real data. For instance, exponential smoothing algorithms give preference to recent trends by exponentially fading off the weights of past factors.

B. *Streaming Operation*

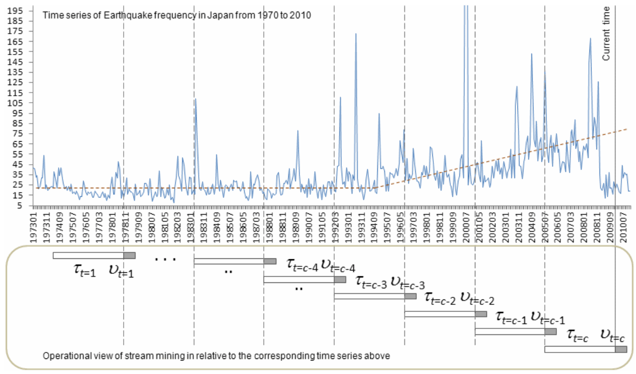
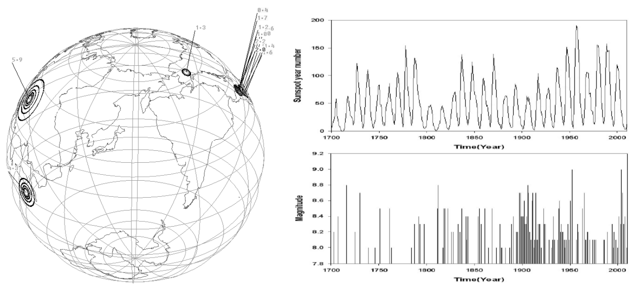


Figure 2. Operational view of adaptive forecasting process – the time series is being forecasted in the fashion of sliding-window.

As shown in Figure 2 above, the adaptive forecasting process operates in the form of data streaming that could be briefly described in the following sequence of steps:

1. The time series is arriving in segments each of which has a certain length, $|\tau|$ = size of observation period. The length is a user-defined variable that should be calibrated in advance. When the length is too long, the forecast will be as good as using a global trend and time lags would exist in between each pair of successive forecasts. If the length is too short, the forecasting accuracy degrades due to insufficient amount of past data were used. In regression, some suggested the minimum length to be the number of cycles equals to one plus the number of coefficients.
2. With each segment of time series data under observation, extract a comprehensive set of salient characteristics (to be explained in the next section). While most of the characteristics would be used for training up the forecasting techniques selector, some useful salient features can be used for other data analysis such as outlier detection, intra-correlation of earthquake events, and inter-correlation between earthquakes and other external activities. Some examples are shown in Figure 3.



(a) (b)

Figure 3. Examples of other analysis by using the statistical characterstics of Earthquake time series: (a) Intra-correlation between earthquake events that happen at the same time across different locations of the Earth, (b) comparison of earthquakes of magnitute > 8 and the number of sun spots.

3. A real-time decision tree classifier which is implemented by Hoeffding Tree [21] algorithm updates itself with the given salient characteristics and the feedback errors from the previous evaluation. The decision tree chooses a suitable forecasting method automatically upon the arrival of a new data segment. The fundamental reason for using Hoeffding Tree is its real-time ability in handling high-speed data stream; its decision tree model can be updated (refreshed) almost instantly whenever a piece of fresh data arrives. In contrast, traditional decision tree requires reading up the whole set of past records plus the update for retraining the model.
4. Once a forecasting method is chosen, it predicts the future points by computing over the observation points within $|\tau|$. $|\upsilon|$ is the length of forecast-ahead period; once again, it is a predefined variable by user. In the example shown in Figure 2 which consists of a time series of earthquake frequencies occurred in Japan from 1973 to 2010, υ/τ is 0.1 where υ=6 months for testing and τ=66 months for training. So the ratio is approximately 0.1 that is common in data mining for training and testing a decision tree. Generally it is known that the accuracy will drop as the forecast-ahead period increases and large observation points are needed for a far-ahead prediction.

Without exhaustive mathematical proofs, streaming operation for adaptive forecasting has its advantage over traditional forecasting from a global trend by looking at the example in Figure 2 intuitively. Represented by a dotted line along with the time series, the linear regression trend is divided into two distinctive eras. Prior to year 1993 the time series, as shown by the real data of earthquake frequencies that took place in Japan obtained from USGS, exhibits a relatively flat gradient. The frequency of earthquake, however, started to climb almost steadily from 1993 onwards. Consequently this observation is indicating a fact that the trend of increase in earthquake frequency suddenly picked up in 1993. And this fact is telling us that the use of trend information (as a salient feature) must be adaptive in our forecasting method. That is, the trend information should be observed periodically and only the most updated information akin to local trend should be used for forecasting. As a counter-example if the global trend that embraces the series from 1973 to 2010 is used for forecasting future events from now on, the predicted values would likely be underestimated because there was indeed an up shift in the trend in 1993.

The adaptive forecasting process ensures the characteristics of the time series are always up-to-date, and the appropriate forecasting method is used based on the current values of the salient characteristics. This fundamentally changes the forecasting paradigm or tradition, from examining the whole series by manually chosen forecasting methods, to a new dynamic and adaptive streaming manner that facilities real-time and accurate forecast.

C. *Decomposing the Time Series into Salient Features*

Salient characteristics in the context of time series forecasting are generally known to be statistical features that spur out beyond their nearby data points. Visually they are the patterns that are noticeable, outstanding, and prominent or can be just easily identified in relative to other parts of the time series. Descriptive statistics are usually the mechanisms for generating these salient characteristics from the time series data. In the adaptive forecasting methodology, decomposing the salient features from the time series is a task of Obtain Information. The salient features as components of a time series can be extracted include but not limited to those listed below:

TABLE I. SOME USEFUL SALIENT FEATURES FOR OUR MODEL

Salient features	Description
Level	Average value of the time series, (mean, variance)
Trend	A reasonably long-term sweep or general direction of movement in a time series
Seasonality	Short-term cyclical behavior that can be observed several times
Noise	Random variation
Outlier	Outstanding values, known as white noise
Auto-correlation	Cross-correlation of a time series with itself, measured in autocovariance
Stationarity	Measure of a stochastic process by its change of joint probability distribution. Sometimes measured by stationary transition probabilities
Random walk	A phenomenon that a time series changes by the same probability distribution and the movements are independent of each other. In this case the past movement of a data point fails to predict its future movement. The data points take a random and unpredictable path.

In addition to those listed above which were used in our experiment, there are many other potential salient features derived from the time series that may relate to selection of the best forecasting method: residuals from previous prediction, causal forces, consistent trends, contrary series, damped trends, decay forces, decomposition, discontinuities, extrapolation, growth forces, inconsistent trends, instabilities, opposing forces, regressing forces, reinforcing series, start-up series, and supporting forces etc. There are also other tests and procedures like Anderson Darling normality test, information derived from Box pierce test, histogram, and QQ plot which shall be exploited in this project. Other factors that might influence the accuracy of the final forecasting result pertain to the streaming operation of our model, such as the choices of τ, the observation period and υ, the prediction period ahead.

According to [22] it is essential to know the salient characteristics of the time series prior to choosing a suitable forecasting method, especially for exponential smoothing method that largely based on the trend and seasonality of the past data points to predict the future ones. In the work [22] it was already shown that the exponential smoothing forecasts can become more accurate, if the better the diagnosis of the salient characteristics was done on the time series.

D. Decision Tree As Forecaster Selector

Extended from Box-Jenkins methodology which finds suitable types of autoregressive (AR) and moving average (MA) techniques from the salient features of the past values of the time series to make forecasts, our model flexibly covers a wider range of techniques. The underlying assumption for the adaptive model is that the time series is composed of a mix of salient features which shall not be forecasted very accurately by any single model alone, and these salient features change their values across time. Therefore a flexible scheme is needed which does not only select the best forecaster automatically, the selection scheme can have learning ability that dynamically fine-tune the 'decision rules' by itself as time goes.

In order to support such adaptive forecaster selector, a light-weight decision tree based on Chernoff-Hoeffding bounds from stream-mining [23] is adopted as a core engine for implementation. Readers who want to obtain details about the decision tree are referred to [23]. The design of the decision tree is quite standard, default parameters are used. The main advantage of the so-called light weight decision tree is its adaptive ability to readjust its structure (rules as tree branches) when input data stream is feeding in. In other words, the classifier is able to adapt to the changing salient characteristics of the time series as new samples of time series are coming in, and it will always choose the relative suitable forecasting method for the new samples.

The rules in the decision trees need to change though at the start the basic tree structure is predefined (pre-trained) according to the experiences learnt from our experiments. The branching conditions at the tree nodes are more or less floppy in a sense that the values are numeric and should be adjusted by the learning experience in the form of error (residual) feedbacks. For example, IF Trend=*strong*, THEN Regression_Model is chosen; How strong then is *strong*? We started the decision tree with default values, and let the data and forecast errors do the fine-tuning – the rules will be periodically updated by reassessing the performance.

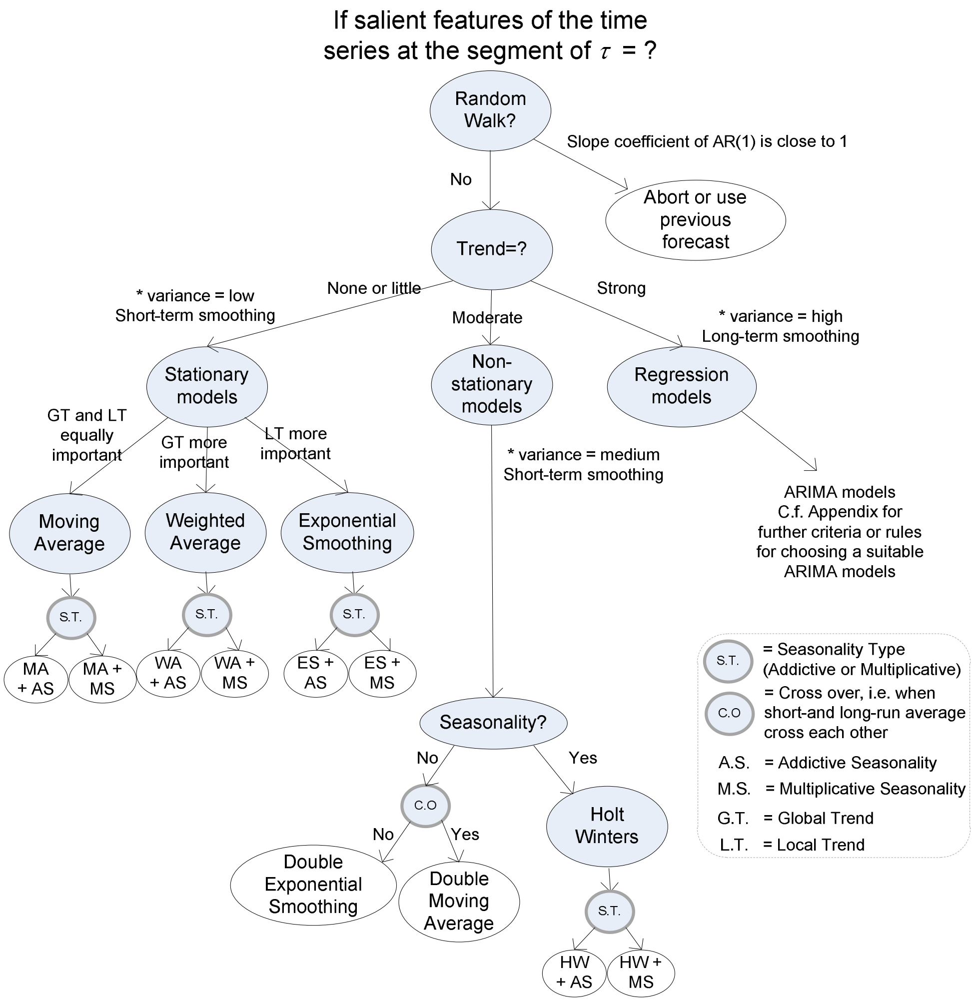


Figure 4. Decision tree for deciding which forecasting method to use, based on the salient characteristics of the time series that are currently under observation.

As can be seen from the decision tree in Figure 4, the very first thing to determine whether the segment of time series is predictable in a sense that its past can predict its future or not. Such predictability of the series can be evaluated by checking the behavior of random walk. A random is a series in which changes from one time period to the next are completely random. It is a special case of AR(1) model in which the slope coefficient is equal to 1. In our experiment of Japan earthquake frequency time series, the AR(1) model has a slope coefficient of 0.5722 and standard error 0.0383 that is more than 10 standard errors away from 1, signifying that the series is not a random walk. (In contrast a financial series of S&P500 monthly closing prices between May 1995 and August 2003 has a slope coefficient of 0.985 and a standard error 0.015 [24]).

Since the predictability of our earthquake data is validated, the decision tree then proceeds to select a suitable forecasting method. From experiences, as shown in our decision tree, trends, stationarity and seasonality of a series are significant factors for choosing a prediction method. In general the higher the node or the conditions for the branching split is the more decisive power it has in the decision making. Nevertheless the example in Figure 4 shows a decision tree that is constructed with conditions of the main salient features. There are other salient factors, perhaps slightly less important are not shown but they are there to further extend out the tree with more conditions down at the branches.

IV. Experiments

Forecasting experiments are run by using our proposed adaptive model and other well-known forecasting methods alone, in order to validate the efficacy of the new model.

The time series data we used were real data originally downloaded from USGS archive. The raw data are historical records of earthquakes that happened in Japan from January 1973 to December 2010 of magnitudes of all levels. For the sake of general interest, the data are transformed into frequency of earthquake per month, with Richter magnitude of at least scale 3. The data has a mean frequency 36.09 per month, standard deviation 33.22, maximum 482 and minimum 8 times per month respectively. By just visually inspecting the time series as shown in Figure 2, the data has a complex mix of salient characteristics; it is somewhat stationary with a slow shift of a rising trend, obviously at the pivot year of 1993, and a sharp drop at the end of the year 2008. A new trend seems to develop from 2008 afterwards, and apparently this local trend is quite unknown due to the shortage of data between 2008 and current year (2011). The seasonality of the time series, as shown in the ACF Plot of autocorrelations in Figure 5 has a set of values exceeding the UCI and LCI, indicating that there is a substantial extent of seasonality though not very strong. The seasonal cycles lack of a strong regularity too as seen in the plot. At the decision tree this salient characteristics should fall upon Holt-Winter model and occasionally into autoregression models as the applied methods. The PACF Plot of partial autocorrelations suggests that the seasonal data depend on other external factors most of the time as well.

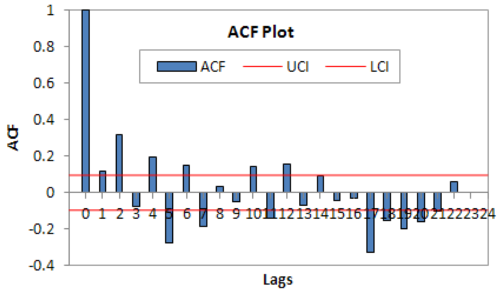


Figure 5. ACF Plot of the earthquake frequency time series

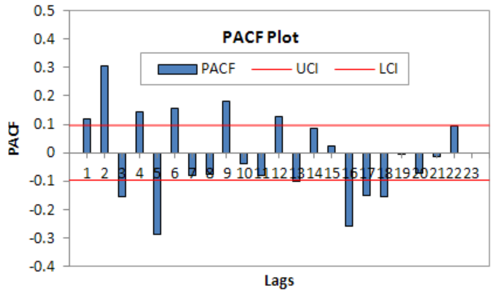


Figure 6. PACF Plot of the earthquake frequency time series

Preliminary results were obtained from running the proposed adaptive model based on the decision tree specified in Figure 4. In general the adaptive model shows its advantage in finding a suitable model therefore producing supposedly the best available result. As observed from the preliminary results from Table 2, the adaptive model can achieve a slightly higher accuracy than most of the other methods which are applied alone, except when there are large fluctuations and sudden change in the patterns. This suggests that additional salient features other than those shown in Figure 4 in the decision tree ought to be used.

TABLE II. COMPARISON OF ADAPTIVE MODEL AND OTHERS ALONE

Forecast type	Average accuracy over all the segments
Holt-Winter	81.40%
Addictive HW	87.20%
Multiplicative HW	80.20%
ARIMA	70.40%
Moving Average	76.70%
Adaptive Model	85.10%

Our adaptive model was probably impaired by the mis-classification by the trend strength into ARIMA. ARIMA was under-performed in time series forecasting in this case because the trend was not clearly strong and the trend might have badly affected by the outlier. A snapshot of forecasting performance as shown in Figure 7 sheds some light on the causes. Evidently the forecast by ARIMA under predicted most of the peaks especially the large ones – outliers. The magnitude of under prediction was almost by half in year 2000 July where Japan had an exceptional large number of earthquakes (482 times). Also the model showed a consistently long period of over-prediction throughout year 2010, because of the sudden and sharp change of trend.

V. Conclusion and Future Works

In previous studies combined forecasts have generally been shown to outperform the forecast from the single best model. It is generally known to many researchers that no single forecasting method can always yield the best results. Therefore selection of the appropriate forecasting methods for different time series has been a popular research topic. Most of the works in the past however concentrated on defining certain fixed rules or guidelines in choosing the best forecasting method. The full time series was often considered under this forecaster selector by diagnosing its salient characteristics. This automatic selection of forecasting methods was extended and pondered on in this paper, with the aim of designing an adaptive model that can be applied in real-time forecasting while the time series was analyzed on the fly. A dynamic decision tree from stream mining was applied in the new model which is able to adjust the rules as reflected by the structure of decision tree. The time series is analyzed in consecutive segments, each of which the salient characters are extracted and used for the decision tree to decide on the forecasting method. An experiment was conducted by taking the live earthquake data from Japan. Importantly this particular time series shows a good example of a complex series that embrace of different salient characteristics at different times; therefore, intuitively no single method should be used to forecast the whole time series since different forecasting method has its strength and weakness for different values of salient characteristics.

The contribution of this paper is a framework and methodology that can be programmed in an automated system for continuous forecasting, such as earthquake pre-warning system, for instance.

References

[1] Yaohui Bai, Jiancheng Sun, Jianguo Luo, and Xiaobin Zhang, "Forecasting financial time series with ensemble learning", 2010 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2010, pp. 1-4.

[2] Ying Jiang, Yicheng Ye, and Qin Wang, "Study on Weighting Function of Weighted Time Series Forecasting Model in the Safety System", 2011 Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2011, pp. 1-4.

[3] Xiao Shengling, Wang Wei, and Wang Bo, "Study on forecasting method of highway port cargo volume", 2010 International Conference on Logistics Systems and Intelligent Management, Volume 3, 2010, pp. 1831-1834.

[4] Shen Fu-Ke, Zhang Wei, and Chang Pan, "An Engineering Approach to Prediction of Network Traffic Based on Time-Series Model", JCAI 2009 International Joint Conference on Artificial Intelligence, 2009, pp. 432-435.

[5] Kawaguchi, K., Kaneda, Y., and Araki, E., "The DONET: A real-time seafloor research infrastructure for the precise earthquake and tsunami monitoring", OCEANS 2008 - MTS/IEEE Kobe Techno-Ocean, 2008, pp. 1-4.

[6] Ji Chang-Peng and Liu li-li, "Real-time Detection for Anomaly Data in Microseismic Monitoring System", International Conference on Computational Intelligence and Natural Computing, (CINC 2009), Volume 2, 2009, pp. 307-310.

[7] Makridakis S., Hibon M., "Accuracy of forecasting : an empirical investigation with discussion", Journal of the Royal Statistical Society A, Volume 142, 1979, pp. 97-145.

[8] Makridakis S., Chatfield C., Hibon M., Lawrence M., Mills T., Ord K., and Simmons L. F., "The M2 competition : a real-time judgmentally based forecasting study", International Journal of Forecasting, Volume 9, 1993, pp. 5-23.

[9] Makridakis S. and Hibon M., "The M3-Competition: Results, Conclusions and Implications", International Journal of Forecasting, volume 16, 2000, pp. 451-476.

[10] Hyndman R. and Khandakar Y., "Automatic Time Series Forecasting: The forecast Package for R", Journal of Statistical Software, Volume 27, Issue 3, July 2008, pp. 1-22.

[11] Lemke C. and Gabrys B., "On the benefit of using time series features for choosing a forecasting method", Proceedings of the 2nd European Symposium on Time Series Prediction, Porvoo, Finland, 2008.

[12] Ragsdale C. and Plane DR., "On modeling time series data using spreadsheets", Omega 2000, The International Journal of Management Science, Elsevier, 2000, Volume 28, Issue 2, pp. 215–21.

[13] Rasmussen R., "On time series data and optimal parameters", Omega 2004, The International Journal of Management Science, Elsevier, 2004, Volume 32, pp. 111–120.

[14] Gelper S., Fried R. and Croux C., "Robust forecasting with exponential and Holt-Winters smoothing", Technical report KBI 0718, Faculty of Economics and Applied Economics, Katholieke Universiteit, Leuven, April 2007, pp. 1-22.

[15] Engle, R. F., Granger, C.W. J., and Hallman, J. J., "Merging short-and long-run forecasts: An application of seasonal co-integration to monthly electricity sales forecasting", Journal of Econometrics, Volume 40, 1989, pp. 45-62.

[16] Casals, J., Jerez, and M., Sotoca, S., "Modelling and forecasting time series sampled at different frequencies", Journal of Forecasting, Volume 28, 2009, pp. 316-342.

[17] Granger, C. W. J., "Implications of seeing economic variables through an aggregation window", Ricerche Economiche, Volume 47, 1993, pp. 269-279.

[18] Armstrong, J. S. (Ed.) (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston: Kluwer.

[19] Yaffee, R. A. (2000). Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS. San Diego: Academic Press.

[20] Newbold P. and Bos T., "On exponential smoothing and the assumption of deterministic trend plus white noise data-generating models", International Journal of Forecasting, Volume 5, Issue 4, Elsevier Science, 1989, pp. 523-527.

[21] Fong S. and Yang H., "Enabling Real-time Business Intelligence by Stream Data Mining, Book Chapter of "Data Mining", Kimito Funatsu (Ed.), ISBN: 978-953-307-547-1, Intech, 05 January 2011, Vienna, Austria.

[22] Fomby T., "Exponential Smoothing Models", Technical report, Department of Economics, Southern Methodist University, June 2008, pp. 1-23.

[23] Yang H. and Fong S., "Investigating the Impact of Bursty Traffic on Hoeffding Tree Algorithm in Stream Mining over Internet", The 2nd International Conference on Evolving Internet (INTERNET 2010), September 2010, Valencia, Spain, pp. 147-152.

[24] Shmueli G., Patel N., and Bruce P., "Data Mining for Business Intelligence", Wiley, 2nd Edition, 2010, pp. 332.

[25] Arsham H., "Time-Critical Decision Making for Business Administration", Online: http://home.ubalt.edu/ntsbarsh/stat-data/Forecast.htm, 2003, [last access: 1/8/2011].

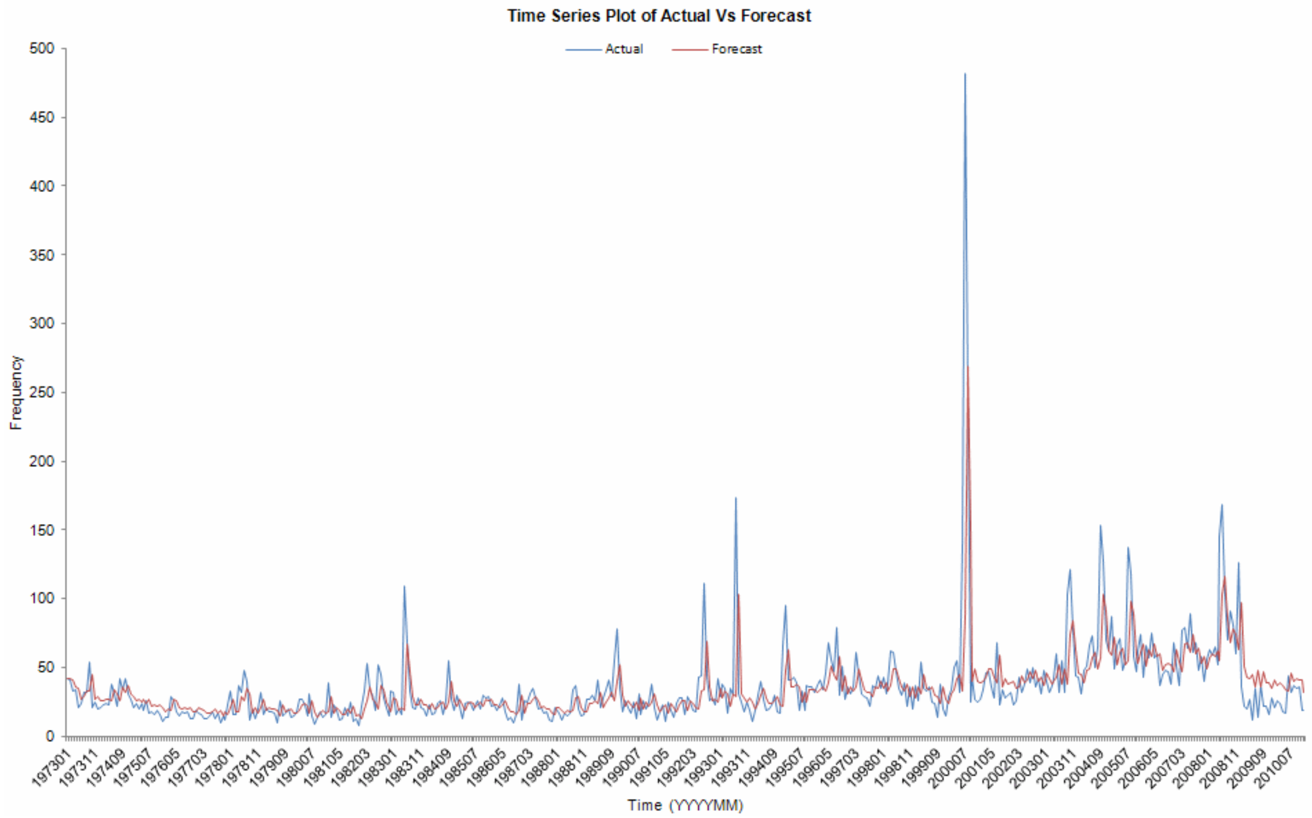


Figure 7. Time series plot of earthquake frequncies that took place in Japan from 1973 to 2010; the forecasting method used is ARMIA(1,1,1)

Appendix - Autoregressive Model Selection Criteria (Source: [25])

1. If none of the simple autocorrelations is significantly different from zero, the series is essentially a random number or white-noise series, which is not amenable to autoregressive modeling.

2. If the simple autocorrelations decrease linearly, passing through zero to become negative, or if the simple autocorrelations exhibit a wave-like cyclical pattern, passing through zero several times, the series is not stationary; it must be differenced one or more times before it may be modeled with an autoregressive process.

3. If the simple autocorrelations exhibit seasonality; i.e., there are autocorrelation peaks every dozen or so (in monthly data) lags, the series is not stationary; it must be differenced with a gap approximately equal to the seasonal interval before further modeling.

4. If the simple autocorrelations decrease exponentially but approach zero gradually, while the partial autocorrelations are significantly non-zero through some small number of lags beyond which they are not significantly different from zero, the series should be modeled with an autoregressive process.

5. If the partial autocorrelations decrease exponentially but approach zero gradually, while the simple autocorrelations are significantly non-zero through some small number of lags beyond which they are not significantly different from zero, the series should be modeled with a moving average process.

6. If the partial and simple autocorrelations both converge upon zero for successively longer lags, but neither actually reaches zero after any particular lag, the series may be modeled by a combination of autoregressive and moving average process.