A HYBRID APPROACH BASED ON ARIMA AND ARTIFICIAL NEURAL NETWORKS FOR CRIME SERIES FORECASTING

MOHD SUHAIMI MOHD ZAKI

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

SEPTEMBER 2014

This dissertation is dedicated to my late father Mohd Zaki bin Hasan, my mother Rusni binti Siking and my sisters Suhaila, Suhanim and Suhanom for their endless support and encouragement.

ACKNOWLEDGEMENT

First and foremost, I would like to express heartfelt gratitude to my supervisor **Dr. Mohamad Shukor bin Talib** for his constant support during my study at UTM. He inspired me greatly to work in this project. His willingness to motivate me contributed tremendously to our project. I have learned a lot from him and I am fortunate to have him as my mentor and supervisor.

Besides, I would like to thank the authority of Universiti Teknologi Malaysia (UTM) for providing me with a good environment and facilities such as Computer laboratory to complete this project with software which I need during process.

ABSTRACT

Crime forecasting is an interesting application area of research with ARIMA and ANN models offer a good technique for predicting time series. Time series data often contain both linear and nonlinear patterns. Therefore, neither ARIMA nor neural networks can be adequate in modeling and predicting time series data. In this study, a hybrid ARIMA and neural network model is proposed to predict crime series data. The hybrid approach for the crime series prediction is tested using 216-month observations of four crime category that are Non-Domestic Violence Related Assault, Break and Enter Non Dwelling, Steal from Retail Store and Steal from Person. Specifically, the results from the hybrid model provide a good modeling framework capable of capturing the nonlinear nature of the complex time series and thus producing more accurate predictions. The accuracy results from the hybrid models for the four case studies are 92.08%, 91.78%, 93.62 and 94.13%, respectively, which are satisfactory in common model applications. Predicted crime data from the hybrid model are compared with those from the ARIMA and neural network using the performance measures. As the result, the hybrid model provides a better accuracy over the ARIMA and neural network models for crime series forecasting.

ABSTRAK

Peramalan jenayah merupakan bidang kajian yang sangat menarik dengan teknik pemodelan ARIMA dan rangkaian neural menawarkan penyelesaian yang baik dalam meramalkan siri masa. Data siri masa lazimnya bercorak lurus dan tidak lurus. Oleh yang demikian ARIMA dan rangkaian neural masing-masing tidak berkeupayaan untuk memodel dan meramal data siri masa secara bersendiri. Di dalam kajian ini, model gabungan di antara ARIMA dan rangkaian neural dicadangkan untuk meramal data siri jenayah. Pemodelan secara gabungan ini diuji terhadap empat set data yang masing-masing dikategorikan sebagai Serangan Berkaitan Keganasan Bukan Kediaman, Pecah Masuk Kediaman Mewah, Mencuri dari Kedai Runcit dan Mencuri dari Seseorang yang kesemua set tersebut mengandungi 216 data. Secara khususnya hasil yang diperolehi daripada gabungan tersebut menunjukkan kerangka pemodelan yang dibangunkan itu boleh dipercayai dan berupaya mengenalpasti kerumitan corak tidak lurus data siri masa yang seterusnya dapat menjana ramalan yang lebih tepat. Peratus ketepatan ramalan yang diperolehi daripada pendekatan gabungan tersebut bagi keempat-empat kajian kes adalah 92.08%, 91.78%, 93.62% dan 94.13% di mana peratusan yang terhasil itu cukup memuaskan berdasarkan penilaian biasa. Data ramalan yang terhasil daripada pemodelan gabungan telah dibuat perbandingan dengan yang terhasil daripada pemodelan ARIMA dan rangkaian neural dengan menggunakan beberapa pengukur prestasi. Secara keseluruhannya, pendekatan gabungan telah menunjukkan prestasi peramalan yang lebih baik berbanding ARIMA dan rangkaian neural untuk peramalan siri jenayah di dalam kajian yang dibuat ini.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SYMBOL	TITLE
ACF	Autocorrelation Function
ANN	Artificial Neural Network
ARCH	Autoregressive Conditional Heteroscedatic
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BEND	Break and Enter Non Dwelling
Lr	Learning rate
Mc	Momentum rate
MSE	Mean Squared Error
NARX	Non Linear Autoregressive Neural Network with External Input
NDVRA	Non-Domestic Violence Related Assault
PACF	Partial Autocorrelation Function
SFP	Steal from Person
SFRS	Steal from Retail Store
SVR	Support Vector Regression
TAR	Threshold Autoregressive

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Forecasting is a process of predicting or estimating the future events by referring to historical data. The information provided about the potential future events and their consequences is able to make important decisions confidently. Forecasting can be carried out by using the availability of time series data. The type of data is time-oriented or a sequence of observations regarding to the variable of interest. This kind of forecasting is known as time series forecasting.

Time series forecasting is an active domain of research that has become increasingly important in various fields of research, such as business, economics, finance, science and engineering. With time series forecasting, the data that consists historical observations are analyzed in a way to develop an appropriate model which describes the inherent structure of the series. It is obvious that a successful time series forecasting depends on an appropriate model fitting. This model is then used to generate future values for the series. A lot of efforts have been spent over the decades in developing efficient models to improve the forecasting accuracy. As the result, currently various important time series forecasting models have been evolved.

Study in crime has its importance. Obviously crime is an omnipresent challenge to societies. The study of Harrendorf et al. (2010) gives some broad comparable international crime cases in homicide, assault, rape, robbery, burglary, motor vehicle theft and kidnapping but generally crime can be divided into two major categories which are violent crime and property crime. Ehrlich and Saito (2010) examine various aspects of criminal activity such as its determinants, different types of crime, policy responses and methods to face criminal activity. One of the methods is time series forecasting and it will be conducted in this study. The ability to predict can serve as a valuable source of knowledge for law enforcement agencies, both from tactical as well strategic perspectives. Forecasting can help a police department's performance by strategic deployment efforts and efficient investigation direction.

1.2 Problem Background

In recent times, there is the growing manifestation among stake holders that crime cannot be controlled exclusively through the action of the police and criminal justice administrators. Always the primary target of the police had been the persons and their criminality, for example, examining the modus operandi of serial criminals, and arresting them (Gorr and Harries, 2003). A motivating factor in the shift of focus from the offender to the offence is the reality that brought the idea of new forms of approach on crime prevention against the backdrop of the apparent failure of the police, courts and prisons to stem the rising crime rates in many societies.

In spite of many limitations of criminal statistics in the current societies, data generated from the case gathered are used in dealing with the problem of crime both in developed and developing countries. For example Australia, the modern trends of criminological studies are that police, executives, judiciaries, prison administrators, parole authorities, social workers and researchers in the various multi disciplinary which subjects relating to crime and crime control spend resources in terms of money and time in assembling, quantifying and analyzing criminality. Empirical data on crime rate serves as a veritable source of information on the rate of crimes in the society. The implications of these data or specifically time series data will enable us to know the pattern and trend of crimes in the country and further assist us to be able to plan for the prevention and curtailment of crime. Crime forecasting is of recent development. As a follow up to the successful crime mapping in the 1990s, the US National Institute for Justice (NIJ) awarded some grants to study crime forecasting for police as an extension of crime mapping with the objective forecasting crime one period ahead (Gorr and Harries, 2003).

In term of forecasting, there are two types of time series modeling known as linear and nonlinear models. Currently there are no single models which able to model both linear and nonlinear pattern of time series data. There are many studies shown that nonlinear models such as ANN are significantly better in forecast than the conventional linear models. ARIMA is one of the conventional linear models and mostly used in time series forecasting (Shahwan and Odening, 2007). But the major limitation of these models is the pre-assumed linear form of the associated time series which becomes inadequate in many practical situations. ARIMA model assumed that in a time series the future values have a linear relationship with current and past values as well as with white noise, thus estimations by ARIMA model may not be adequate for nonlinear time series that inherited a complex real-world problems that often nonlinear (Chen, et al., 2005).

Nevertheless, with a major progress in model combination or hybrid has given a way out in improving the forecasting performance. Both theoretical and empirical findings agreed that by combining different models together is effectively improve the forecasting performance, particularly when the hybrid is a combination of different models from each other (Berardi and Zhang, 2003). Many different techniques have been combined and proposed in order to overcome the deficiencies of a single model and hoping to produce more accurate results. Hybrid models can either be homogeneous, such as using differently configured neural networks or heterogeneous, such as using ARIMA and ANN models (Balkin and Ford, 2000).

Furthermore, a hybridization of ARIMA and ANN models as linear and nonlinear model is extensively studied by researchers since it produces promising results (Alwee, Shamsuddin and Sallehuddin, 2013). Further discussions about the literature of the hybrid between ARIMA and ANN are given in Section 2.6. In conducting this study, we are using a medium size of data and tried to control the training results from over fitting. Many factors will surely influenced the results from selected architecture and parameters of ANN model but the motivation behind this hybrid approach is largely due to the fact that the crime rates is often complex to predict. The selection of architecture and parameters of ANN model is made due to the limitation of study period and exploration efforts of this new domain personally.

1.3 Research Objective

The objectives of the research are:

- To obtain a forecasting model by using ARIMA.
- To obtain a forecasting model by using ANN with the correspond residuals as input data into the model.
- To implement a hybrid approach by using the results from ARIMA and ANN.

1.4 Research Scope

The delimitation of the research noted as the following:

- The dataset been used is a univariate historical time series.
- The actual data from the dataset are presented by month, offense type and area.
- The dataset is obtained from NSW Police of Australia website.

1.5 Thesis Outline

The organization of this thesis is summarizes as follows. In Chapter 2, the concept of ARIMA, ANN and the hybrid models are reviewed. In Chapter 3, the methodology of each modeling process has been described. In Chapter 4, the results from the methodology are discussed and finally Chapter 5 will conclude the discussions.

REFERENCES

Alain, N. K., 2002. Macroeconomic Forecasting: A Comparison Between Artificial Neural Networks and Econometric Models, published M.Sc. thesis, Rand Afrikaans University.

Armano, G., Marchesi, M. & Murru, A., 2005. A hybrid genetic-neural architecture for stock indexes forecasting, Information Sciences 170, 3–33.

Alwee, R., Shamsuddin, S. M. H. & Sallehuddin, R., 2013. Hybrid Support Vector Regression and Autoregressive Integrated Moving Average Models Improved by Particle Swarm Optimization for Property Crime Rates Forecasting with Economic Indicators, The Scientific World Journal, Vol 2013.

Bates, J. M. & Granger, W. J., 1969. The combination of forecasts, Operation Research 20, 451–468.

Berardi, V. L. & Zhang, G. P., 2003. An empirical investigation of bias and variance in time series forecasting: modeling considerations and error evaluation, IEEE Transactions on Neural Networks 14 (3), 668–679.

Balkin, S. D. & Ord, J. K., 2000. Automatic neural network modeling for univariate time series, International Journal of Forecasting 16, 509–515.

Box, P. & Jenkins, G. M., 1976. Time Series Analysis: Forecasting and Control, Holden-day Inc., San Francisco, CA.

Chakraborty, K., Mehrotra, K., Mohan, C. K., Ranka, S., 1992. Forecasting the behavior of multivariate time series using neural networks, Neural Networks 5, 961–970.

Chakradhara, P. & Narasimhan, V., 2007. Forecasting exchange rate better with artificial neural network, Journal of Policy Modeling 29, 227–236.

Chen, K. Y. & Wang, C. H., 2007. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan, Expert Systems with Applications 32, 2007, 254–264.

Chen, Y., Yang, B., Dong, J., Abraham, A., 2005. Time-series forecasting using flexible neural tree model, Information Sciences 174 (3–4), 219–235.

De Gooijer, G. J. & Hyndman, R. J., 2006. 25 years of time series forecasting, International Journal of Forecasting, vol. 22, pp. 443-473.

Ediger, V. & Akar S., 2007. ARIMA forecasting of primary energy demand by fuel in Turkey, Energy Policy 35, 1701–1708.

Ehrlich, I. & Saito, T., 2010. Taxing Guns vs. Taxing Crime: An Application of the “Market for Offenses Model, NBER Working Papers 16009, National Bureau of Economic Research, Inc.

Emdad, K., 2000. Neural Fuzzy Based Intelligent Systems and Applications, In: Lakhmi C.J. and Martin N.M (Eds), Fusion of Neural Networks, Fuzzy Systems, and Genetic Algorithms Industrial Application, The CRC Press International Series on Computational Intelligence, New York, pp.107-139.

Giordano, F., La Rocca, M., Perna, C., 2007. Forecasting nonlinear time series with neural network sieve bootstrap, Computational Statistics and Data Analysis 51, 3871–3884.

Jain, A., Kumar, A. M., 2007. Hybrid neural network models for hydrologic time series forecasting, Applied Soft Computing 7, 585–592.

Karray, F. O., & De Silva, C., 2004. Soft Computing and Intelligent Systems Design: Theory, Tools and Application, 1st edition, England Pearson Education Limited, pp. 223, 2004.

Kim, H., & Shin K., 2007. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets, Applied Soft Computing 7, 569–576.

Khashei, M., Bijari, M., & Raissi, G. H. A., 2009. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs), Neurocomputing 72, 956–967.

Khashei, M., Hejazi, S. R. & Bijari, M., 2008. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting, Fuzzy Sets and Systems 159, 769–786.

Nahmias, S. W., 2005. Production and Operations analysis, McGraw-Hill/Irwin, New York.

Pai, P. F. & Lin, C. S. 2005. A hybrid ARIMA and support vector machines model in stock price forecasting, Omega 33, 497–505.

Sharkey, A., 2002. Types of multinet system Proceedings of the Third International Workshop on Multiple Classifier Systems Lecture Notes in Computer Science, 2364, Springer, London, 2002, 108–117.

Shaikh A., 2004. Primer on using Neural Networks for Forecasting Market Variables, [Accessed January, 4, 2008].

Shahwan, T. & Odening, M., 2007. Forecasting Agricultural Commodity Prices using Hybrid Neural Networks. In S.-H. Chen, P. Wang &T.-W. Kuo (Eds.), *Computational Intelligence in Economics and Finance* (pp. 63-74): Springer Berlin Heidelberg.

Shi, J., Guo, J., & Zheng, S., 2012. Evaluation of hybrid forecasting approaches for wind speed and power generation time series, Department of Industrial and Manufacturing Engineering, North Dakota State University, USA.

Stergious, C., & Siganos, D., 2009. Neural Networks, www. doc.ic.ac.uk~nd/ surprise_96/journal/vol4/cs11/report.html [Access online, March 2010].

Taskaya, T. & Casey, M. C., 2005. A comparative study of autoregressive neural network hybrids, Neural Networks 18, 781–789.

Tseng, F. M., Yu, H. C., & Tzeng, G. H., 2002. Combining neural network model with seasonal time series ARIMA model, Technological Forecasting & Social Change 69, 71–87.

Tabachnick, B.G. & Fidell, L.S., 2001. Using multivariate statistics, 4th ed., Person Education Company, USA.

Yu, L., Wang, S., & Lai, K. K., 2005. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates, Computers and Operations Research 32, 2523–2541.

Zhang, G. P., 2003. Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50, 2003, 159–175.

Zhou, Z. J. & Hu, C. H., 2008. An effective hybrid approach based on grey and ARMA for forecasting gyro drift, Chaos, Solitons and Fractals 35, 525–529.