Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (1): 167-172, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.167.172

Crime Forecasting using ARIMA Model and Fuzzy Alpha-cut

[1]Noor Maizura Mohamad Noor, [1]Astari Retnowardhani, [2]Mohd Lazim Abd and [1]Md Yazid Mohd Saman
[1]Department of Computer Science, [2]Department of Mathematics, Faculty of Science and Technology,
University Malaysia Terengganu, 21030 Kuala, Terengganu, Malaysia

Abstract: Crime forecasting is one of research in forecasting world. Many forecasting methods have been applied in this field as Naïve lag, exponential smoothing, decomposition method and ARIMA model. This study proposed crime forecasting using Autoregressive Integrated Moving Average (ARIMA) model and fuzzy alpha-cut method. This combination is expected to generate more accurate forecasting result with minimum error. The results will help the decision maker's in making the right decision in crime prevention strategies.

Key words: ARIMA model, crime forecasting, fuzzy alpha-cut

INTRODUCTION

Crime prevention is one of the important roles of the police system in any country. One of the components of crime prevention is crime rate predictions or in other word is crime forecasting. Police will require crime forecasting to make operational and tactical strategies in the future, as like to allocate police patrols in the right area, install CCTV in the right place and plan other operations. A common practice is to identify hot spots in the preceding period based on their geographical location and assume these hot spots will persist into the next period (Pieterse, 2006).

Some researches in crime forecasting have been done by several researchers like Cohen and Gorr (2005), Deadman (2003) and Gorr *et al*. (2003). Many methods have been applied in crime forecasting research. Used of Naïve lag, exponential smoothing methods and classical decomposition by Gorr *et al*. (2003). Use of the decomposition method made by Jiang and Barricarte (2011), Deadman (2003), Chen *et al*. (2008) and Noor *et al*. (2011b) used ARIMA model.

Explored with the use of ARIMA model for crime forecasting has not been widely used. One of the research used ARIMA model discussed by Chen *et al*. (2008). They used the ARIMA model for making short-term forecasting of property crime for one city of China, then compared forecasting results with the Simple Exponential Smoothing (SES) and Holt's two-parameter exponential smoothing (HES). The conclusion of the research is the ARIMA model fits the series better than SES and HES and makes higher accurate forecasting than the other two models. ARIMA model provides forecasting result with the interval range of forecasting value. The interval range describe the possible condition.

To improve interval forecasting accuracy of crime forecasting this study proposes combined between ARIMA model and Fuzzy alpha-cut method for crime forecasting process. The results could provide information about crime trends, especially the possible worst and better condition in the future. This information can help the police in making decision for operational and tactical strategies in crime prevention.

CRIME PREVENTION IN MALAYSIA

In the recent years, volumes of crime had brought serious problems, not least in Malaysia (Menteri, 2010). Such reported in "The Roadmap 2010", the overall index crime rate in Malaysia increases from 746 reported crimes per 100,000 people in 2006 to 767 in 2007, a rise of nearly 3% (Menteri, 2010). The Malaysian government has taken a step forward in "Reducing Crime" program included in Government Transformation Program (GTP). The government target in GTP 2010 in terms of reducing crime is achieved 5% reduction in index crime in overall reported index crime every year and 20% reduction in street crime. As of 31 December 2010, the target of reducing incidents of index crime by 5% and street crime by 20% have been exceeded. For index crime, there has been a reduction of 32,297 cases (15%) while Street Crime has dropped by 13,193 cases (35%). Even though the index crime rate has dropped by 15%, there are still 177,520 cases reported as of December 2010. This shows crime is still occurring (Menteri, 2011a).

The Royal Police Malaysia (RMP) in recent years has several operational and tactical strategies in crime prevention. The operations and strategy are separation of pedestrian walkways with railings, improved lightings,

Corresponding Author: Noor Maizura Mohamad Noor, Department of Computer Science, Faculty of Science and Technology, University Malaysia Terengganu, 21030 Kuala, Terengganu, Malaysia

mirrors, safety alarms, GIS-based (geographic information system) crime mapping and closed-circuit televisions or CCTVs (implemented separately under the supervision of the Ministry of Housing and Local Government). At the center of this crime prevention effort is a Safe City Monitoring System (SCMS) for monitoring the current crime at that time (Menteri, 2011b). To support the government programme, require crime forecasting system to help operational and tactical strategy. By knowing the crime rate prediction in the future it is expected that police can be more accurate in determining the strategy choices, for example are allocated police patrols and install CCTV in the selected hot spot area.

To fulfill the need we propose Crime Forecasting System. In forecasting process of this system we used ARIMA model and fuzzy alpha-cut.

MATERIAL AND METHODS

This research is a part of the framework Crime Management System (CeMaS), shown in Fig. 1, where this framework proposed for police to support reducing crime program (Noor *et al*., 2011a).

Discussion in this paper is focused on Crime Forecasting Module. Our proposed system consists of 3 parts, shown in Fig. 2. There is database, operation, output. Part 1 is crime database which contain about historical crime data, location and crime forecasting data. Part 2 describe two process. The first process is a forecasting process used ARIMA model and fuzzy alpha-cut method. The second process is crime information retrieval process. The results of crime information retrieval process would be displayed to the decision maker, discussed in Part 3. Wherein the information would be showed trends of crime with visualization on graph, table and mapping.

Crime forecasting process in this study used ARIMA model and Fuzzy alpha-cut method. With combination between ARIMA model and Fuzzy alpha-cut method is expected to generate more accurate interval forecasting results with minimum error.

Autoregressive integrated moving average (ARIMA): In forecasting process, we use ARIMA model. The ARIMA model is denoted by ARIMA (p,d,q), where "p" stands for the order of the auto regressive process, 'd' is the order of the data stationary and 'q' is the order of the moving average process. In ARIMA(p, d, q) model, the future value of a variable is assumed to be a linear function of several past observations and random errors. That is, the underlying process that generates the time series has the form:

$$y_t = \theta_0+\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+...+\varphi_p y_{t-p} +\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-...-\theta_q\varepsilon_{t-q} \quad (1)$$

where, y_t and ε_t are the actual value and random error at time period t, respectively ϕ_i (i = 1,2,...,q) and ϕ_j (i = 1,2,...,q) are model parameters. p and q are integers and often referred to as orders of the model. Random errors ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2 (Zhang, 2003).

The forecasting process used Box Jenkin's methodology for ARIMA model. The Box-Jenkins methodology includes three iterative steps (Khashei *et al*., 2009):

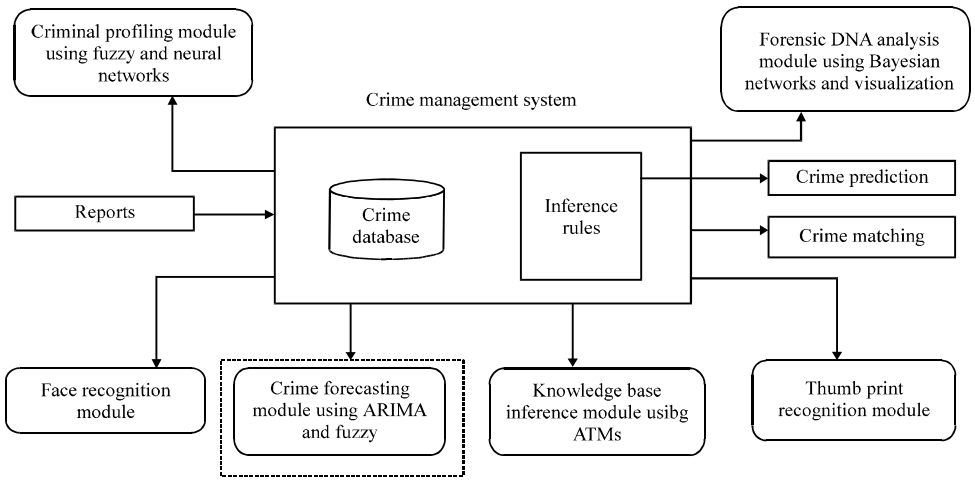

Fig. 1: Crime management system (CeMaS)

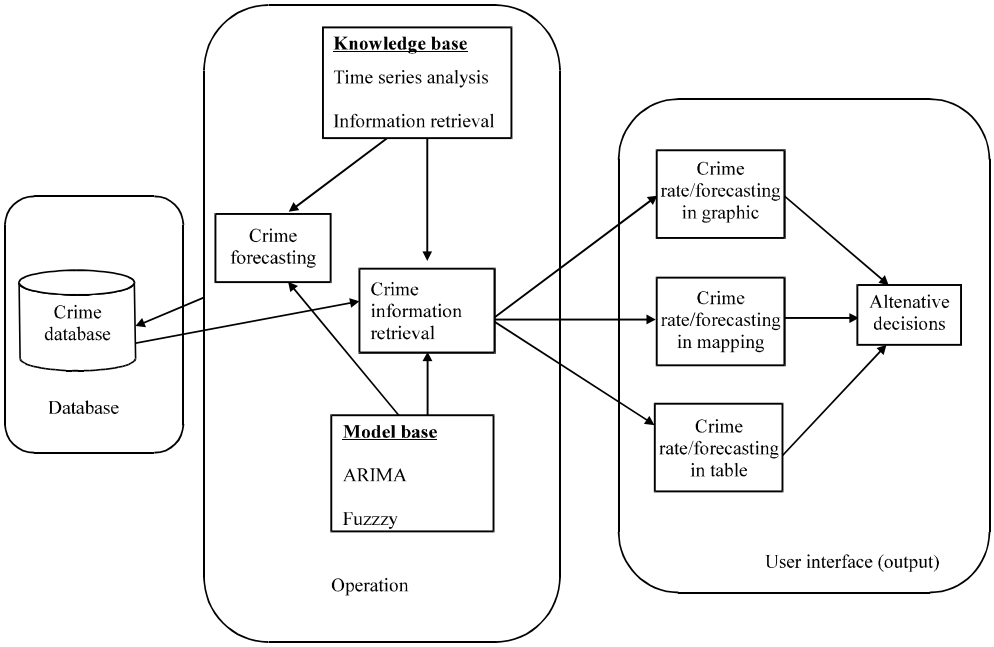

Fig. 2: Crime forecasting module

- Model identification
- Parameter estimation
- Diagnostic checking

Before any model is identified we must consider the pattern of data. After getting the pattern of data we follow the first step Box-Jenkins methodology, that is identification model. In this step, data transformation is often required to make the time series stationary. Stationarity is a necessary condition in building an ARIMA model used for forecasting. In second step several models could be obtained. After having the tentative models the next step is checking the model to obtain the best model. This three-step model building process is typically repeated several times until a satisfactory model is finally selected. Furthermore, the best model can be used for forecasting and the results of forecasting will be stored in a database. Implementation of this process discussed by Noor *et al.* (2011b). Figure 3 describes the process of crime forecasting with Box-Jenkins methodology used historical crime index data.

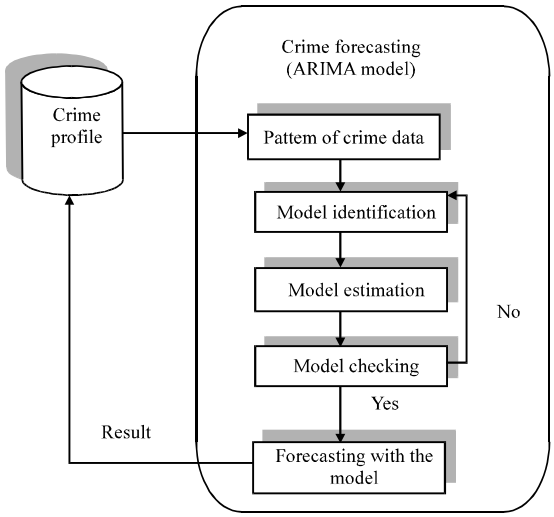

Fig. 3: The crime forecasting process used ARIMA model

Fuzzy alpha-cut (FAC): Alpha cuts are simply threshold levels that convert a fuzzy set into a crisp set. The process of converting a fuzzy set to a crisp one is called defuzzification. An alpha-cut A of a fuzzy number A is defined as the set $\{x \in R|A(x) \geq \alpha\}$. A is completely determined by the collection $(A_\alpha)\alpha \in [0,1]$. An alpha cut is the degree of sensitivity of the system to the behavior under observation. At some point, as the information value diminishes, one no longer wants to be "bothered" by the data. In many systems, due to the inherent limitations of the mechanisms of observation, the information becomes suspect below a certain level of reliability (Kumar and Schuhmacher, 2005).

The membership function is cut horizontally at a finite number of α-levels between 0 and 1. For each

α-level of the parameter, the model is run to determine the minimum and maximum possible values of the output. This information is then directly used to construct the corresponding fuzziness (membership function) of the output which is used as a measure of uncertainty. If the output is monotonic with respect to the dependent fuzzy variable, the process is rather simple since only two simulations will be enough for each α-level (one for each boundary). Otherwise, optimization routines have to be carried out to determine the minimum and maximum values of the output for each α-level. Figure 4 shows an illustration of the alpha-cut of Triangular Fuzzy Number (TFN), where Lo and Up are the lower and upper bounds of the closed interval.

A triangular fuzzy number Ã can be written as Eq. 2 can be defined by a triplet (a,b,c). The membership function $\mu_{\tilde{A}}(x)$ is defined as (Kaufman and Gupta, 1991):

$$\mu_{\tilde{A}}(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x\le b,\\ \dfrac{c-x}{c-b}, & b\le x\le c,\\ 0, & x>c\end{cases} \quad (2)$$

Alternatively, by defining the interval of confidence at level α, we can characterize the triangular fuzzy number as (Cheng, 1999):

$$\forall\alpha\in[0,1],\tilde{A}^{\alpha}=[Lo\ ,Up\ ]=[(b-a)\alpha+a,-(c-b)\alpha+c] \quad (3)$$

Standard deviation (SD): Standard Deviation (SD) shows how much variation or "dispersion" exists from the average (mean, or expected value). A low SD indicates that the data points are spread out over a large range of values. Equation 4 shows formula of the SD of the sample:

$$S=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \quad (4)$$

RESULTS AND DISCUSSION

The model proposed is combination between ARIMA model and fuzzy alpha-cut method, as shown in Fig. 5. This combination is aimed to obtain more accurate interval forecasting result.

This process has six steps. The first step is making the pattern of the data and the next steps are Box Jenkin's methodology steps. The fifth step is forecast with the best model ARIMA. The results of the ARIMA forecasting models based on actual data will be converted to fuzzy observations with lower and upper membership function. The arithmetic operation on the fuzzy alpha-cut, for instance the alpha-cut method is applied to the data through the fuzzy environmental process.

Implementation of first step until fifth step has been made on Noor *et al*. (2011b). The result found ARIMA (0,1,1) is the best model in this study with of mean square error (MSE) = 7411. Then crime forecasting is done using this model. The forecasting result is shown in Table 1.

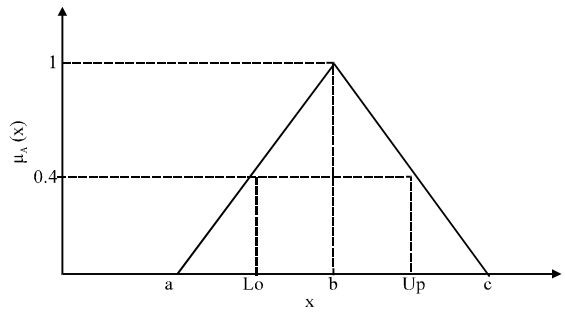

Fig. 4: Fuzzy numbers Ã with α-cuts

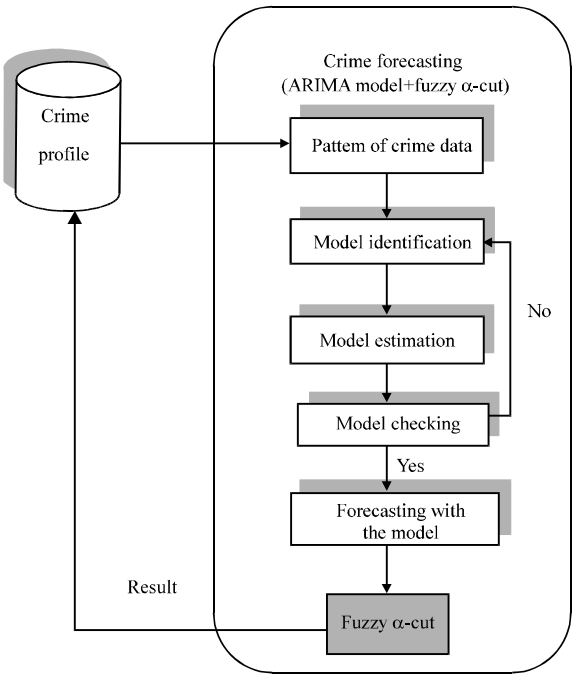

Fig. 5: Crime forecasting process with fuzzy α-cut

Table 1: Forecasting results with ARIMA (0,1,1)

Period (month)	Forecast	Lower	Upper	Actual
31	1132.14	963.37	1300.91	1182
32	1119.91	950.61	1289.21	1130
33	1107.67	937.84	1277.50	951
34	1095.44	925.08	1265.80	1215
35	1083.20	912.32	1254.09	1083
36	1070.97	899.56	1242.38	1185

Table 2: Alpha cut results

	α = 0.2		α = 0.5	
Period (month)	Lower	Upper	Lower	Upper
31	997.12	1267.150	1047.75	1216.52
32	984.47	1255.350	1035.26	1204.56
33	971.80	1243.530	1022.75	1192.58
34	959.15	1231.728	1010.26	1180.62
35	946.49	1219.910	997.76	1168.64
36	933.84	1208.090	985.26	1156.67

Table 3: Standard deviation results

	Without α-cut		α = 0.2		α = 0.5	
ORI forecasting	Lower	Upper	Min.	Max.	Min.	Max.
22.88878743	23.87605048	21.89992093	23.67859784	22.09769419	23.38241891	22.39435412

The forecasting values with lower and upper values became a crisp value by a triplet (a,b,c). For example used values from month 31, the lower value as a, upper value as c and forecast value as b. Then the value of a = 963.37, b = 1132.14 and c = 1300.91.

For the next step with Eq. 3 we calculated TFN with crisp values from forecasting values and α-cut values 0.2 and 0.5. The result of calculation became Lo and Up bounds of the closed interval. The α-cut result shows in Table 2. Standard Deviation (SD) performed after the results of α-cut obtained. The SD results show in Table 3.

Table 2 shows the interval values with α = 0. 5 closer to the forecast value, better than interval values with α = 0. 2.

Table 3 shows the values of SD from lower and upper forecasting results, lower and upper with α = 0.2 and with α = 0.5. The better result with the lowest SD is lower and upper with α = 0.5. The SD with α = 0.5 nearest to the SD forecasting.

CONCLUSION

Two values of α-cut presented for getting the better interval values of crime forecasting which closer to forecasting values. More closer to the forecasting value indicate more closely to actual value. The SD results with α = 0.5 is lower than SD with α = 0.2. This is indicated that SD with α = 0.5 is better than α = 0.2. The proposed study in combining fuzzy α-cut with ARIMA is expected to the accuracy of interval forecasting results of the crime, especially for minimum and maximum values of forecasting. Further study is needed to strengthen the proposed combination methods. Used of various α-cut values for testing combination methods will be present the better conclusion.

ACKNOWLEDGMENTS

The authors would like to thank for continuous supports given by Datuk Seri Mohd Bakri bin Mohd Zinin, Director of Criminal Investigation Department, Royal Police Malaysia (RPM).

This study was supported by a grant from Fundamental Research Grant Scheme (FRGS) Ministry of Higher Education (MOHE) Malaysia, with vot. number 59201.

REFERENCES

Chen, P., H. Yuan and X. Shu, 2008. Forecasting crime using the ARIMA model. Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery, Volume 5, October 18-20, 2008, Jinan, Shandong, China, pp: 627-630.

Cheng, C.H., 1999. Evaluating weapon systems using ranking fuzzy numbers. Fuzzy Sets Syst., 107: 25-35.

Cohen, J. and W.L. Gorr, 2005. Development of Crime Forecasting and Mapping Systems for Use by Police. H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, USA., Pages: 150.

Deadman, D., 2003. Forecasting residential burglary. Int. J. Forecasting, 19: 567-578.

Gorr, W., A. Olligschlaegerb and Y. Thompson, 2003. Short-term forecasting of crime. Int. J. Forecasting, 19: 579-594.

Jiang, Q. and J.J.S. Barricarte, 2011. A crime rate forecast and decomposition method. Int. J. Criminol. Sociol. Theory, 4: 648-656.

Kaufman, A. and M.M. Gupta, 1991. Introduction to the Fuzzy Arithmetic: Theory and Applications. Van Nostrand Reinhold Co., New York, USA., ISBN-13: 9780442008994, Pages: 361.

Khashei, M., M. Bijari and G.A.R. Ardali, 2009. Improvement of auto-regressive integrated moving average models using fuzzy logic and Artificial Neural Networks (ANNs). Neurocomputing, 72: 956-967.

Kumar, V. and M. Schuhmacher, 2005. Fuzzy uncertainty analysis in system modelling. Comput. Aided Chem. Eng., 20: 391-396.

Menteri, J.P., 2010. Reducing Crime. In: Government Transformation Programme: The Roadmap, Menteri, J.P., (Ed.). Chapter 6, Unit Pengurusan Prestasi Dan Perlaksanaan, Malaysia.

Menteri, J.P., 2011a. Government transformation programme. Annual Report 2010. www.pmo.gov.my/dokumenattached/GTP_AR2010_ENG_FINAL.pdf

Menteri, M.J.P., 2011b. Safe city monitoring system: A Malaysia first in crime prevention. Performance Management and Delivery Unit. http://www.pemandu.gov z.my/gtp/Press_Releases-@-Safe_City_Monitoring_System.aspx

Noor, N.M.M., A.F. Ghazali, M.Y.M. Saman and Z. Zafarina, 2011a. Supporting decision-making for forensic DNA analysis in crime investigation using visualization. Proceedings of the 2nd European Conference for the Applied Mathematics and Informatics, December 29-31, 2011, Montreux, Switzerland, pp: 36-41.

Noor, N.M.M., R. Astari, M. Norizan, M.S.M. Yazid and A.M. Lazim, 2011b. A framework of decision support for crime forecasting in Malaysia. Proceedings of the 2nd European Conference for the Applied Mathematics and Informatics, December 29-31, 2011, Montreux, Switzerland, pp: 31-35.

Pieterse, C., 2006. Predicting Criminal Incidents to Support Police Planning. Universiteit van Amsterdam, Amsterdam .

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