

# 行政院國家科學委員會專題研究計畫 成果報告

## 應用 DNA 計算於預測技術探討 研究成果報告(精簡版)

計畫類別：個別型  
計畫編號：NSC 95-2416-H-004-040-  
執行期間：95年08月01日至96年07月31日  
執行單位：國立政治大學應用數學學系

計畫主持人：吳柏林

計畫參與人員：碩士班研究生-兼任助理：陳孝煒、高健維

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫涉及專利或其他智慧財產權，1年後可公開查詢

中華民國 96 年 08 月 28 日

行政院國家科學委員會補助專題研究計畫  成果報告  期中進度報

## DNA 計算及在預測技術之應用

計畫類別： 個別型計畫  整合型計畫

計畫編號：NSC 95-2416-H-004-040

執行期間：95 年 8 月 1 日至 96 年 7 月 31 日

計畫主持人：吳柏林

共同主持人：

計畫參與人員：

成果報告類型(依經費核定清單規定繳交)： 精簡報告  完整報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、  
列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年  二年後可公開查詢

執行單位：國立政治大學應用數學學系

中華民國 96 年 8 月 20 日

## Introduction

The DNA computing in the forecasting process is applied for several reasons: First, a problem specific computing will be easier to design and implement, with less need for model/functional complexity and flexibility. Secondly, the types of soft computational problems that DNA based computing may be able to effectively solve social competence for economic conditions that a dedicated processor would be naturally reasonable. As well, these problems will be likely to require extensive time that would preclude the need for a more versatile and interactive system that may be able to be implemented with a universal computing machine.

The practical possibility of using DNA molecules as a medium for computation was first demonstrated by Adleman [2]. He demonstrated a proof-of-concept use of DNA as form of computation which was used to solve the seven-point Hamiltonian path problem [1], [3]. The primary intention of Adelman's work was to prove the feasibility of molecular computation, and also gave an indication that the emergence of this new computational paradigm could provide an advantage over conventional electronic computing techniques. Specifically, DNA has shown to have massively parallel processing capabilities which allow a DNA based computer to solve complicated computational problems in a reasonable amount of time.

The main advantage offered by most proposed models of DNA based computation is the ability to handle millions of operations in parallel. The massively parallel processing capabilities of DNA computers may give them the potential to find tractable solutions to otherwise intractable problems, as well as potentially speeding up large, but otherwise solvable, polynomial time problems requiring relatively few operations. Many different models exist within the paradigm of "classical" DNA computation [4], [8], [10], [15], and [20] each of them with different advantages and degrees of applicability to classes of problems.

On the other hand, in the field of humanity and social science, fuzzy statistics and fuzzy forecasting have attracted many attentions lately. This is a nature result because of the complicated phenomenon of humanity and society which is hard to be fully explained by traditional models. Taking stock market as an example, the essence of closing price is uncertain and indistinctive. Moreover, there are many factors could influence the closing price, such as trading volume and exchange rate, etc. Therefore, if we merely consider the closing price of previous day to construct our forecasting model, we are capable of estimating the future trend, rather than unexpected or unnecessary loss.

Upon applying fuzzy logic in the time series analysis, the first step is to identify how to integrate linguistic variable analysis methods in solving the autoregressive relation problem of the dynamic data. Chiang et al. [6] presented self-learning methods to modify fuzzy models for dynamic system in linguistic field. Later, Huarng [10] proposed a fuzzy linguistic summary as one of the data mining function to discover useful knowledge from database. In fact, fuzzy relation equations are easier to be understood and applied than decision tables or decision rules.

In view of this, many researchers have adopted fuzzy relation equations for time series analysis and forecasting. For instance, Wu and Hung [26] proposed a fuzzy identification procedure for ARCH and Bilinear models. Kumar and Wu [12] used fuzzy statistical techniques in change period's detection of nonlinear time series. Chen and Hwang [5] proposed the two-factors time-variant fuzzy time series model and developed two algorithms for temperature prediction. Huarng [10] and Tseng et al. [24] proposed heuristic models by integrating problem-specific heuristic knowledge with Chen's [5] model to improve forecasting.

In this research, we propose a fuzzy time series modeling process with DNA computing. This method is applied to the financial time series data, and then forecast future trend while comparing the forecasting performance. From the empirical studies, it is shown that our proposed method demonstrates an appropriate and efficient performance of prediction for exchange rates.

## Encoding Scheme

Next will be a series of biochemical process. Assume we have a time series data with  $n$  records, which means there exist  $n-1$  first order difference. A number of  $n-1$  test tubes are set up in represent each first order difference  $\Delta X_t$ . Two DNA sequences are designed in represent the "up" trend and "down" trend, denote as DNA1 and DNA2, respectively. These two DNA sequences are identical beside its own special designed PCR primers

denote as UF, UR, DF, and DR (UF: DNA1 forward primer; UR: DNA1 reverse primer, DF: DNA2 forward primer, DR: DNA2 reverse primer) as shown in Fig. 3.1.

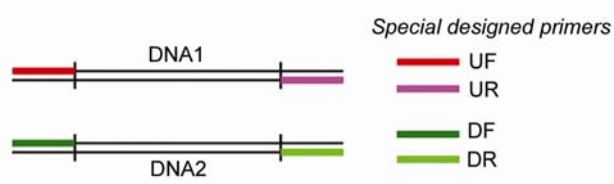


Fig 3.1 DNA encoding scheme. DNA1 and DNA2 are designed in represent “up” trend and “down” trend respectively. DNA1 and DNA2 have its own special design primers.

Synthesizing short single-stranded DNA is now a routine process. The molecules can be made by an auto-programming machine called DNA synthesizer. Itineraries can then be produced from the encodings by linking them together in proper order. To accomplish this we can take advantage of the fact that DNA hybridizes with its complementary sequence. Random itineraries can be made by mixing difference encodings.

The ratio of up trend primers (UF, UR) and down trend primers (DF, DR) for each test tube is set up corresponding with  $\Delta X_t/T$  according to the Table 3.2.

Table 3.2 The ratio of primer for each test tube

$\Delta X_t/T$	[-1.0,-0.9]	(-0.9,-0.6]	(-0.6,-0.3]	(-0.3,-0.1]	(-0.1,0.1]	(0.1,0.3]	(0.3,0.6]	(0.6,0.9]	(0.9,1.0]
$P_{up}:P_{down}$	9:1	8:2	7:3	6:4	5:5	4:6	3:7	2:8	1:9

## Making Prediction

Finally, according to the quantity of DNA1 and DNA2 retrieved from previous step, we define the ratio of up-and-down trend as the auto-correlation coefficient by:

$$U_n = \frac{\sum_{i=2}^n DNA1_i}{\sum_{i=2}^n DNA1_i + \sum_{i=2}^n DNA2_i} \quad (3.1)$$

$$D_n = \frac{\sum_{i=2}^n DNA2_i}{\sum_{i=2}^n DNA1_i + \sum_{i=2}^n DNA2_i} \quad (3.2)$$

In the fuzzy forecasting for one step,  $(U-D)T$  is transferred into fuzzy number according to Table 3.1. Hence, the one step fuzzy forecasting becomes:

$$FX_{n+1} = X_n \oplus \left( \frac{m_i}{L_i} + \frac{m_{i+1}}{L_{i+1}} \right) T \quad (3.8)$$

where  $\left( \frac{m_i}{L_i} + \frac{m_{i+1}}{L_{i+1}} \right) T$  means the memberships of the variation with respect to the linguistic variables  $L_i, L_{i+1}$

based on the total variation  $T$ , and  $\oplus$  stands for the fuzzy addition.

The proposed DNA computing algorithm for forecasting is summarized as following:

*Step 1.* For time series  $\{X_t\}$ , decide the total variation of  $\{\Delta X_t\}$  and linguistic variables  $\{L_1, L_2, \dots, L_5\}$  of  $\Delta X_t$ .

*Step 2.* Design the memberships of fuzzy time series w.r.t the linguistic variables

- Step 3. Determine the ratio of primer
- Step 4. Design the DNA sequences and primers
- Step 5. Molecular reaction [PCR]
- Step 6. DNA quantification [Real-time PCR]
- Step 7. Calculate the forecasting value

## Data Analysis

The monthly exchange rates of *JPY/USD* and *EUR/USD* are taken as an example in this study. The exchange rates between 2004/1 and 2005/12 as shown in Fig. 4.1 are retrieved from the Central Bank of China. According to the data source,  $T=\max|\Delta X_t|=5.27$  for *JPY/USD*, and  $T=\max|\Delta X_t|=0.034$  for *EUR/USD* can be obtained.



Figure 4.1 Trend for Exchange Rate: JPY/USD and EUR/USD

## Fuzzy Model Construction

After fuzzifying these data of monthly exchange rate, the method mentioned in Section 3.1 is applied for calculating the data's corresponding memberships in linguistic variables  $L_i$  as illustrated in Tables 4.1 and 4.2.

Table 4.1 The ratios of primer and memberships for JPY/USD

Month	$\Delta X_t / T$	Ratio of primer $P_{up} : P_{down}$	Memberships				
			plunge	drop	draw	soar	surge
2004/1-2	0	5:5	0	0	1	0	0
2004/2-3	0.39	3:7	0	0	0.22	0.78	0
2004/3-4	-0.26	6:4	0	0.5	0.5	0	0
2004/4-5	1.00	1:9	0	0	0	0	1
2004/5-6	-0.58	7:3	0.16	0.84	0	0	0
2004/6-7	-0.02	5:5	0	0.04	0.96	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2005/9-10	0.71	2:8	0	0	0	0.59	0.41
2005/10-11	0.68	2:8	0	0	0	0.64	0.36
2005/11-12	0	5:5	0	0	1	0	0

## Forecasting and Performance

The memberships of *JPY/USD* and *EUR/USD* monthly variations between January 2004 and December 2005 can be observed from Tables 4.1 and 4.2. Finally, the result from DNA forecasting is shown at Tables 4.3 and 4.4. Due to the purpose of this research is to explore the qualitative trend of time series, the memberships are

transformed based on fuzzy rule from the concentration of DNA by  $C = \frac{50 - DNA1\%}{50}$  as illustrated at Table

4.5, which is generated from Table 3.1. The corresponding linguistic variables are obtained to facilitate analysis, and the result is compared with the best ARIMA model  $(p,d,q)$  in Tables 4.6 and 4.7.

Table 4.3 The ratios of two DNAs and memberships for JPY/USD after reaction

Month	Ratio DNA1:DNA2	Memberships				
		plunge	drop	draw	soar	surge
2004/1-2	52:48	0	0.08	0.92	0	0
2004/2-3	25:75	0	0	0	1	0
2004/3-4	69:31	0	0.76	0.24	0	0
2004/4-5	6:94	0	0	0	0.24	0.76
2004/5-6	58:42	0	0.32	0.68	0	0
2004/6-7	54:46	0	0.16	0.84	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2005/9-10	52:48	0	0.08	0.92	0	0
Average	41:59	0	0	0.36	0.64	0
*2005/11	38:62	0	0	0.52	0.48	0
*2005/12	45:55	0	0.0	0.80	0.20	0
*2006/1	47:53	0	0	0.88	0.12	0
*2006/2	42:58	0	0	0.68	0.32	0

\*Note that the second column of Table 4.3 is a pre-experimental result; a more complete experiment should be conducted for the business application.

The MAFA is defined in Section 3.6 for measuring the accuracy of forecasting methods. From Tables 4.6 and 4.7, we can find the  $DNA(1)$  model has better forecasting performance than ARIMA model. The MAFA result is shown at Table 4.8.

The major reason why the prediction cannot hit the real value is that we only consider the greatest membership and omit others memberships. Therefore, only with reasonable forecasting model can we decide investment strategy from forecasting results. Otherwise, without the direction of clear outlines, investors will face a plight as to which information they should take.

Table 4.8 Forecasting accuracy of Best ARIMA and DNA computing from Mean Absolute Forecasting Accuracy

Exchange Currency	Best ARIMA 1,0,1	DNA computing
JPY/USD	81.25%	81.25%
EUR/USD	75%	81.25%

## Concluding Remarks and Future Works

DNA computing is so exciting because of the collaboration of chemists, biologists, mathematicians, and computer scientists to understand and simulate fundamental biological processes and algorithms taking place within cells. Although DNA computing might not replace conventional computing in the near future, they still have endless potentials for other applications. The DNA computing has clear advantages over conventional computing when applied to problems that can be divided into separate, non-sequential tasks. The reason is that for non-decomposable problems, those that require many sequential operations are much more efficient on a conventional computer due to the length of time required to conduct the biochemical operations.

In this research, we tried to make an appropriate process of constructing fuzzy time series model and use this model to forecast the exchange rate of  $JPY/USD$  and  $EUR/USD$ . Compare the DNA forecasting and DNA fuzzy forecasting model with traditional ARIMA model by the performance of forecasting accuracy, we can find that the DNA model has better forecasting performance than that of traditional ARIMA model. We hope this method will provide a new forecasting technique for investors to make optimal decision with fuzzy information.

In spite of the forecasting performance for DNA forecasting model, there are some problems for further studies. For example:

- (i) To make a general rule for fuzzy order identification instead of the Markov relation. DNA computing can be more accurately described as a collection of new computing paradigms rather than a single focus. Each of these different paradigms within molecular computing can be associated with different potential applications that may prove to place them at an advantage over conventional methods. Many of these models share certain features that lend them to categorization by these potential advantages. However, there exists enough similarities and congruencies that hybrid models will be possible, and that advances made in both "classic" and "natural" areas.
- (ii) To extend our result to the multivariate fuzzy time series case. In fact, how to solve the nonstationary or seasonal factors for the time series are still open questions.
- (iii) In this research, we adopt five-ranking classification and transform the time series data into fuzzy numbers through membership functions. However, seven-ranking classification used in social sciences may be used in future studies for special situation, and it is yet to prove where it will provide significant improvement on forecasting performance?
- (iv) Future applications might make use of the error rates and instability of DNA based computation methods as a means of simulating and predicting the emergent behavior of complex systems. This could pertain to weather forecasting, economics, and lead to more a scientific analysis of social science and the humanities. Such a system might rely on inducing increased error rates and mutation through exposure to radiation and deliberately inefficient encoding schemes. Similarly, methods of DNA computing might serve as the most obvious medium for use of evolutionary programming for applications in design or expert systems. DNA computing might also serve as a medium to implement a true fuzzy logic system.

# 95-96 出席國際學術會議心得報告

2007/4/1

計畫編號	NSC 2416H004040
計畫名稱	應用 DNA 計算於預測技術探討
出國人員姓名	吳柏林
服務機關及職稱	國立政治大學應用數學系教授
會議時間地點	北九州市 Japan
會議名稱	2007 Internation Symposium in Mangement Information System
發表論文題目	Fuzzy Statistical Analysis and Estimation

## 一、參加會議經過

The 2007 Internation Symposium in Mangement Information System , 於 2007 年 3 月 9 日至 11 在北九州市 Japan 舉行, 由日本 Waseda University 主辦, 來自各國之學者專家約有 80 餘人參加, 包括地主國日本外, 台灣, 美國, 加拿大, 韓國, 馬來西亞, 香港, 蘇俄, 捷克等 10 餘國..

本次大會就以下幾項重點主題進行研討

1. Bioinformatics 2. Soft computing 3. Intelligent computing 4. Information mangement 除此之外, 尚有幾個序列的 Workshops 與 Panel discussions 討論有關 DNA computing 等. 日本專家教授在會中特別提到: Intelligent Computing 與 Soft Computing 之生物觀點. Mangement Information and Uncertainty. 以及一些新觀念如: Perception based reasoning, 知識管理與創新計算等新看法。

## 二、與會心得

目前筆者在政大應數研究所開設人工智慧與模糊統計分析, 深感學術研究發展日新月異、一日千里。國科會能給予補助出國出席國際學術研討會, 收穫相當大。希望將來能多利用課餘時間出國做短期研究, 吸收國外新知、及研究方向。回國後繼續開設模糊時間數列分析與預測課程, 指導博、碩士班研究生, 籌辦國際學術研討會, 推動國際經濟與管理學術研究工作。並於國際著名學術論文期刊, 發表學術論文。