ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การพัฒนาระบบช่วยการพยากรณ์แบบออโตรีเกรสซีพโดยใช้วิธีเจเนติกอัลกอริธึม

การสร้างระบบช่วยการพยากรณ์แบบออโตรีเกรสซีพโดยใช้วิธีเจเนติกอัลกอริธึม

โดย

สายสุดา ธนาธรรมนันท์

SAISUDA THANATHAMMANUNT

H004902

ส.ม.
ส663 ก
2550

เลขหมู่.....04902.....
เลขทะเบียน.....
วัน,เดือน,ปี - 6 พ.ย. 2551

b11978764
i.....

อาจารย์ที่ปรึกษา
รศ. ดร. อาริต ธรรมโน

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 2 ปีการศึกษา 2550

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

BUILDING AUTOREGRESSIVE FORECASTING SYSTEM USING GENETIC ALGORITHM

SAISUDA THANATHAMMANUNT

A SYSTEM DEVELOPMENT PROJECT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/ 2007

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2008

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ แสวงหาผลประโยชน์ทางการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การพัฒนาระบบช่วยการพยากรณ์แบบออโตรีเกรสซีพโดยใช้วิธีเจเนติกอัลกอริธึม
นักศึกษา	นางสาวสายสุดา ธนาธรรมนันท์
รหัสนักศึกษา	46066848
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2550
อาจารย์ที่ปรึกษา	รศ. ดร. อาริต ธรรมโน

บทคัดย่อ

การพยากรณ์แบบอนุกรมเวลาโดยวิธีการ Autoregressive เป็นส่วนหนึ่งของการพยากรณ์แบบ Box-Jenkins หรือ ARIMA model ซึ่งเป็นกระบวนการที่ใช้ข้อมูลในอดีตเป็นตัวกำหนดรูปแบบของ Autoregressive model ที่จะนำมาใช้ทำนายข้อมูลนั้นในอนาคต โดยจะได้เป็นสมการในการพยากรณ์ข้อมูล ซึ่งวิธีการที่จะได้มาซึ่ง รูปแบบของ Autoregressive model นั้น มีหลายวิธี ซึ่งไม่มีรูปแบบที่ตายตัวแน่นอน โดยบางวิธีจะอาศัยความสัมพันธ์ในตัวของข้อมูลเองมาวิเคราะห์รูปแบบที่น่าจะเป็นในการทำนาย

โครงงานนี้ นำเสนอวิธีการหารูปแบบของ Autoregressive model โดยใช้ Genetic algorithm เป็นเครื่องมือที่ช่วยในการหารูปแบบที่เหมาะสมที่สุด สำหรับในแต่ละข้อมูลที่นำมาพยากรณ์ โดยเริ่มแรกจะทำการสุ่ม Autoregressive model ขึ้นมาจำนวนหนึ่งเพื่อเป็นประชากรเริ่มต้น และทำการ cross over และ mutation เพื่อให้ได้กลุ่มของ Autoregressive model ในรุ่นถัดไป โดยกระบวนการนี้จะทำไปเรื่อย ๆ เพื่อหารูปแบบที่ดีที่สุดในการพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Building Autoregressive forecasting system using Genetic algorithm
Student	Miss Saisuda Thanathammanunt
Student ID.	46066848
Degree	Master of Science
Program	Information Science
Academic Year	2007
Advisor	Assoc.Prof. Dr.Arit Thammano

ABSTRACT

The Time series forecasting using Autoregressive model is a part of Box-Jenkins method. The autoregressive method get data from the past for create Autoregressive model to forecast this data in the future. There are many methods to generate Autoregressive models. There is possibility to have more than one model and we can not ensure that the model that we choose is best for forecasting.

This project proposes a method to finding the best model using Genetic Algorithm. Genetic algorithm choosing number of models randomly as the first population then, the first population will be crossed over and mutated to generate the next generation population. We repeat this process until getting the best Autoregressive model

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงการฉบับนี้สำเร็จได้อย่างดีด้วยคำแนะนำและคำปรึกษาจาก ผศ. ดร. อาริต ธรรมโน อาจารย์ผู้ควบคุมโครงการที่กรุณาให้คำปรึกษา คำแนะนำต่างๆตลอดจนตรวจสอบและแก้ไข จนทำให้โครงการนี้เสร็จสมบูรณ์

ขอกราบพระคุณคณาจารย์คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า

ขอขอบคุณ คุณกิติศักดิ์ พินิจ และบริษัทพาราแอควานส์อินโฟเทค ที่อำนวยความสะดวกในเรื่องเวลา และทรัพยากรที่ใช้ในการทำโครงการนี้

ขอขอบคุณเพื่อนๆ ที่ให้กำลังใจและให้ความช่วยเหลือมาโดยตลอดนับตั้งแต่เข้ามาศึกษาในสถาบันแห่งนี้

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดามารดาและครอบครัวของข้าพเจ้าที่เป็นกำลังใจและให้การสนับสนุนในทุกเรื่องๆ

สุดท้ายนี้ขอขอบพระคุณทุกๆ ท่านที่ไม่ได้กล่าวนามไว้ ณ ที่นี้ซึ่งได้ให้ความร่วมมือและให้ความช่วยเหลือจนโครงการนี้สำเร็จลุล่วงได้ด้วยดี

สายสุดา ธนาธรรมนันท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อ I

ABSTRACT II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่ 1 1

บทนำ 1

1.1 ความเป็นมาของโครงการ 1

1.2 วัตถุประสงค์ของโครงการ 1

1.3 ขอบเขตของโครงการ 1

1.4 ขั้นตอนและวิธีการศึกษาโครงการ 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 2

1.6 รายละเอียดในบทต่าง ๆ 2

บทที่ 2 3

Time series forecasting 3

2.1 Averaging method 3

2.1.1 Simple averages 3

2.1.2 Moving averages 3

2.2 Exponential smoothing method 5

2.2.1 Single exponential smoothing 5

2.2.2 Holt's linear method 6

2.3 Regression method 7

2.3.1 Simple regression 7

2.3.2 The correlation coefficient 9

2.3.3 Coefficient of determination 10

2.4 ARIMA model 12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.4.1 The autocorrelation function (ACF) 13

2.4.2 The partial autocorrelation function (PACF) 13

2.4.3 การทำข้อมูลให้เป็น stationary 13

2.4.4 Backshift notation 16

2.4.5 การนำ ARIMA model มาใช้กับข้อมูล time series 16

2.4.6 Autoregressive model ลำดับที่ 1 16

2.4.7 Moving average model ลำดับที่ 1 17

2.4.8 Autoregressive model ในลำดับที่สูงขึ้น 18

2.4.9 Moving average model ในลำดับที่สูงขึ้น 19

2.4.10 ARIMA model ในแบบผสม 19

บทที่ 3 21

Genetic Algorithm 21

3.1 หลักการของ Genetic algorithm 21

3.2 ขั้นตอนของ genetic algorithm 23

3.2.1 การคัดเลือกประชากรต้นแบบ 23

3.2.2 กระบวนการ crossover 26

3.2.3 กระบวนการ mutation 27

บทที่ 4 28

การประยุกต์ใช้งาน Autoregressive model ร่วมกับ Genetic Algorithm 28

4.1 การแทนค่าโครโมโซมของ Autoregressive model ใน Generic algorithm 28

4.2 การทำข้อมูลให้เป็น stationary 28

4.3 การทำงานของ Generic algorithm 29

4.1 สุ่มค่าประชากรเริ่มต้น 29

4.2 แทนค่าโครโมโซม 29

4.3 หาค่า fitness และค่า ค่า probability ของแต่ละประชากร 30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.4 สร้าง roulette wheel selection และหาค่า cumulative เพื่อใช้ในการคัดเลือกประชากร 30
4.4 สุ่มคัดเลือกประชากร 31
3.2.4 กระบวนการ crossover 32
3.2.5 กระบวนการ mutation 33
บทที่ 5 34
การพัฒนาระบบและหน้าจอการทำงาน 34
5.1 สภาพแวดล้อมการพัฒนาระบบ 34
5.2 การจัดเตรียมสภาพแวดล้อมในการทำงาน 34
5.3 ขั้นตอนการทำงาน 34
5.3.1 การเตรียมข้อมูล 34
5.3.2 การประมวลผล 37
5.4 การพัฒนาระบบและหน้าจอการทำงาน 37
5.4.1 หน้าจอส่วนที่รับ input และประมวลผล 38
5.4.1 หน้าจอส่วนที่แสดง output 39
5.4 ผลการทดสอบโปรแกรม 40
บทที่ 6 43
สรุปผลการพัฒนาระบบและข้อเสนอแนะ 43
6.1 ผลการวิเคราะห์และออกแบบระบบ 43
6.2 ประโยชน์ที่คาดว่าจะได้รับ 43
6.3 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต 43
บรรณานุกรม 44
ประวัติผู้เขียน 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 การพยากรณ์ค่าของยอดขายโดยใช้วิธี moving averages.....4

2.2 การพยากรณ์ยอดขายโดยใช้วิธี exponential smoothing.....5

2.3 การพยากรณ์ยอดขายโดยใช้วิธี Holt's Linear method.....6

2.4 ยอดการส่งออกสินค้าและราคาสินค้า.....8

2.5 การคำนวณค่า R^2 ของราคาสินค้า (P_i) และยอดการส่งออกสินค้า (S_i) 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 กราฟสมการเส้นตรงจากข้อมูลระหว่างราคาสินค้า กับจำนวนการนำเข้าสินค้า.....9

2.2 ขั้นตอนการพยากรณ์โดยใช้ ARIMA model.....12

2.3 Dow-Jones index.....14

2.4 Dow-Jones index ที่ผ่านการทำ first-order differencing แล้ว.....15

2.5 กราฟ ACF และ PACF ของ ARIMA(1,0,0).....17

2.6 กราฟ ACF และ PACF ของ ARIMA(0,0,1).....18

3.1 ตัวอย่างของโครโมโซมในรูปแบบของเลขฐานสอง.....21

3.2 การ crossover ของโครโมโซม A และ B.....22

3.3 การ mutation ของโครโมโซม C.....22

3.4 Roulette wheel selection.....25

4.1 Roulette wheel selection ที่คำนวณจากค่า probability.....30

5.1 แสดงข้อมูล Dow-Jones index.....34

5.2 แสดงข้อมูล Number of internet users.....34

5.3 Dow-Jones index ที่ผ่านการทำ first-order differencing แล้ว.....35

5.4 Number of internet users ที่ผ่านการทำ first-order differencing แล้ว.....35

5.5 หน้าจอการประมวลผลและแสดงผลของโปรแกรม.....36

5.5 หน้าจอส่วนที่รับ input และประมวลผล.....37

5.7 หน้าจอแสดง output.....38

5.8 แสดงผลของค่า Dow-Jones index ที่พยากรณ์เปรียบเทียบกับค่าจริง.....39

5.9 ค่าความคาดเคลื่อนที่ได้จากการพยากรณ์ค่า Dow-Jones index 40

5.10 แสดงผลของค่า Number of internet users ที่พยากรณ์เปรียบเทียบกับค่าจริง.....41

5.11 ค่าความคาดเคลื่อนที่ได้จากการพยากรณ์ค่า Number of internet users.....41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของโครงการ

ในปัจจุบัน การพยากรณ์ข้อมูลมีส่วนในการช่วยตัดสินใจเกี่ยวกับการดำเนินธุรกิจมากขึ้น ตัวอย่างเช่น การพยากรณ์ยอดขายของสินค้า หรือการพยากรณ์ปริมาณการใช้สินค้าต่าง ๆ ซึ่งสามารถนำวิธีการพยากรณ์ข้อมูลแบบอนุกรมเวลามาใช้ได้ โดยนำข้อมูลในอดีตมาใช้พยากรณ์ข้อมูลในอนาคต ซึ่งมีวิธีการพยากรณ์ได้หลายวิธี โดยโครงงานนี้เราจะนำเอาวิธีการของ Autoregressive ซึ่งเป็นส่วนหนึ่งของวิธีการพยากรณ์แบบ Box-Jenkins หรือ ARIMA model มาใช้ในการพยากรณ์ชุดของข้อมูลที่เราสนใจ

ปัญหาของการพยากรณ์โดยวิธี Autoregressive คือ ไม่มีความแน่นอนในการกำหนด Autoregressive model เพื่อนำมาแปลงเป็นสมการที่ใช้พยากรณ์ข้อมูล ในข้อมูลที่นำมาพยากรณ์ชุดหนึ่ง ๆ นั้นอาจจะมี Autoregressive model ในหลาย ๆ รูปแบบที่มีความเป็นไปได้ในการนำมาใช้งาน ในโครงงานนี้จึงนำเอาวิธีการของ Genetic algorithm มาใช้ในการหา Autoregressive model ที่เหมาะสมที่สุด โดย Genetic algorithm จะทำการคัดเลือก Autoregressive model มาจำนวนหนึ่งเพื่อเข้ามาเป็นประชากรรุ่นแรก และทำการ crossover และ mutation เพื่อทำการคัดเลือก Autoregressive model ที่ได้มาเป็นประชากรรุ่นต่อไป Genetic algorithm จะทำงานในลักษณะแบบนี้ไปเรื่อย ๆ จนได้ Autoregressive model ที่เหมาะสมที่สุดในการใช้งานออกมา

1.2 วัตถุประสงค์ของโครงการ

1. เพื่อศึกษาขั้นตอนและวิธีการพยากรณ์แบบอนุกรมเวลาโดยใช้วิธีการ Autoregressive model
2. ศึกษาหลักการและวิธีการของGenetic algorithm
3. นำความรู้ที่ศึกษาไปพัฒนาโปรแกรมที่ช่วยในการสร้าง Autoregressive model ที่เหมาะสมผ่านกระบวนการ Genetic algorithm

1.3 ขอบเขตของโครงการ

เป็นการพัฒนาระบบเพื่อการพยากรณ์ข้อมูลโดยใช้การทำนายแบบ Autoregressive และนำ Genetic algorithm มาช่วยในการคัดเลือก Autoregressive model ที่เหมาะสมที่สุดในการใช้พยากรณ์ข้อมูล โดยมีขอบเขตการทำงานดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. สามารถพัฒนาโปรแกรมตามทฤษฎีและอัลกอริทึมได้ถูกต้อง
2. สามารถนำไปใช้งานกับข้อมูลทั่วไปได้
3. สามารถนำเสนอผลลัพธ์การทำงานได้ถูกต้อง

1.4 ขั้นตอนและวิธีการศึกษาโครงการ

1. กำหนดหัวข้อ เป้าหมายและวัตถุประสงค์ตลอดจนขอบเขตการทำงานของระบบ
2. ศึกษาทฤษฎี บทความ งานวิจัยและหนังสือที่เกี่ยวข้องกับหลักการของการพยากรณ์แบบ Box-Jenkins และ Genetic algorithm
3. ศึกษาเทคนิคในการพัฒนาโปรแกรม
4. พัฒนาระบบตามที่ได้ศึกษาข้อมูลมา
5. ทำการทดสอบระบบ
6. ปรับปรุงและแก้ไขข้อผิดพลาด
7. สรุปผลการศึกษาและการดำเนินงาน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเข้าใจหลักการและเทคนิคการพยากรณ์อนุกรมเวลาโดยวิธีการ Box-Jenkins
2. เพื่อเข้าใจถึงหลักการและวิธีการของ Genetic algorithm
3. เพื่อนำโมเดลที่พัฒนาไปประยุกต์ใช้กับการพยากรณ์ข้อมูลต่างๆต่อไป

1.6 รายละเอียดในบทต่าง ๆ

บทที่ 1 บทนำ กล่าวถึงความเป็นมาของโครงการ ความมุ่งหมายและวัตถุประสงค์ ขอบเขตของการศึกษาและพัฒนา และขั้นตอนการศึกษา

บทที่ 2 Time series forecasting
นำเสนอเนื้อหาและกระบวนการในการพยากรณ์อนุกรมเวลา

บทที่ 3 Genetic algorithm
นำเสนอเนื้อหาและกระบวนการของ Genetic algorithm

บทที่ 4 การประยุกต์ใช้งาน Autoregressive model ร่วมกับ Genetic Algorithm
นำเสนอการใช้งาน Autoregressive model ร่วมกับ Genetic Algorithm ในการพัฒนาโปรแกรม

บทที่ 5 การพัฒนาระบบและหน้าจอการทำงาน

บทที่ 6 สรุปผลการพัฒนาระบบและข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

Time series forecasting

Time series forecasting หรือ การพยากรณ์ข้อมูลแบบอนุกรมเวลานั้น เป็นวิธีการพยากรณ์ที่นำเอาข้อมูลในอดีตมาใช้พยากรณ์ข้อมูลในอนาคต ซึ่งวิธีการพยากรณ์นั้น สามารถทำได้หลายวิธี เช่น นำข้อมูลในอดีตมาหาค่าเฉลี่ยเพื่อเป็นค่าทำนาย ในอนาคต (Simple averages method) หรือนำข้อมูลมาพล็อตสมการเส้นตรง (Regression method) ซึ่งจะยกตัวอย่างในหัวข้อต่อไปดังนี้

2.1 Averaging method

เป็นการพยากรณ์โดยนำค่าเฉลี่ยของข้อมูลในอดีตมาทำนาย ในที่นี้จะกล่าวถึง 2 วิธีคือ Simple averages และ Moving averages

2.1.1 Simple averages

เป็นการทำนายโดยการนำเอาข้อมูลทั้งหมดในกลุ่มมาหาค่าเฉลี่ยเพื่อนำมาทำนายค่าในคาบเวลาถัดไปโดยมีสมการดังนี้

$$F_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i \qquad (2.1)$$

โดยให้ F_{t+1} เป็น ค่าที่ทำนายได้ในคาบเวลาถัดไป

Y_i เป็น ค่าจริงในคาบเวลา T ใด ๆ

t เป็น จำนวนข้อมูล

2.1.2 Moving averages

เป็นการทำนายโดยจำกัดจำนวนที่จะนำมาหาค่าเฉลี่ยเป็นจำนวน k ข้อมูล โดยข้อมูลที่มีจำนวน t จะเริ่มนำมาหาค่าเฉลี่ยตั้งแต่ ชุดของข้อมูลที่คาบเวลา ที่ t-k+1 จนถึงข้อมูลในคาบเวลาปัจจุบันคือ t

$$F_{t+1} = \frac{1}{k}\sum_{i=t-k+1}^{t} Y_i \qquad (2.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยให้ F_{t+1} เป็น ค่าที่ทำนายได้ในคาบเวลาถัดไป

Y_i เป็น ค่าจริงในคาบเวลา T ใด ๆ

t เป็น จำนวนข้อมูลทั้งหมด

k เป็น จำนวนข้อมูลที่นำมาใช้หาค่าเฉลี่ย

ตัวอย่าง:

แสดงการพยากรณ์ค่าของยอดขายในเดือนธันวาคม จากข้อมูลที่ได้ในแต่ละเดือน โดยใช้วิธี moving averages ในช่วง 3 เดือน และ 5 เดือน

ตารางที่ 2.1 การพยากรณ์ค่าของยอดขายโดยใช้วิธี moving averages

Month	Time period	Observed values	Three-month moving averages	Five-month moving averages
Jan	1	200		
Feb	2	135		
Mar	3	195		
Apr	4	197.5	176.7	
May	5	310	175.8	
Jun	6	175	234.2	207.5
Jul	7	155	227.5	202.5
Aug	8	130	213.3	206.5
Sep	9	220	153.3	193.5
Oct	10	277.5	168.3	198
Nov	11	235	209.2	191.5
Dec	**12**		**244.2**	**203.5**

จากตาราง ค่า moving averages ของเดือนธันวาคม ที่เป็นช่วง 3 และ 5 เดือน จากสมการ จะได้ ค่า k เป็น 3 และ 5 ตามลำดับ โดยแทนค่าสมการได้ดังนี้

Three-month moving averages:

$$F_{12} = \frac{Y_9 + Y_{10} + Y_{11}}{k} = \frac{220 + 277 + 235}{3} = 244.2 \qquad (2.3)$$

Five-month moving averages:

$$F_{12} = \frac{Y_7 + Y_8 + Y_9 + Y_{10} + Y_{11}}{k} = \frac{155 + 130 + 220 + 277 + 235}{3} = 203.5 \qquad (2.4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 Exponential smoothing method

เป็นการพยากรณ์โดยมีการให้ weight ของข้อมูล โดยจะให้ค่า weight ของข้อมูลมีค่ามากที่สุดในข้อมูลปัจจุบันและค่า weight จะลดลงในรูปแบบของ exponential ตามความเก่าของข้อมูล โดยมีหลายวิธีการ เช่น Single exponential smoothing method , Holt's linear method , Holt-Winters' trend and seasonality method

2.2.1 Single exponential smoothing

เป็นการนำค่าที่ทำนายได้ในคาบเวลา F_t มาเปรียบเทียบกับค่าจริงในคาบเวลาเดียวกัน Y_t เพื่อหา error ในการจะนำมาปรับปรุงเพื่อทำนายค่าในอนาคต โดยมีสมการดังนี้

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \qquad (2.5)$$

ให้ α เป็นค่า weight ที่เป็นค่าคงที่ระหว่าง 0 กับ 1

ตัวอย่าง:

แสดงการพยากรณ์ค่าของยอดขายในเดือนธันวาคม โดยวิธีการ exponential smoothing

ตารางที่ 2.2 การพยากรณ์ยอดขายโดยใช้วิธี exponential smoothing

Month	Time period	Observed values	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
Jan	1	200	-	-	-
Feb	2	135	200.0	200.0	200.0
Mar	3	195	193.5	167.5	141.5
Apr	4	197.5	193.7	181.3	189.7
May	5	310	194.0	189.4	196.7
Jun	6	175	205.6	249.7	298.7
Jul	7	155	202.6	212.3	187.4
Aug	8	130	197.8	183.7	158.2
Sep	9	220	191.0	156.8	132.8
Oct	10	277.5	193.9	188.4	211.3
Nov	11	235	202.3	233.0	270.9
Dec	**12**		**205.6**	**234.0**	**238.6**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 Holt's linear method

เป็นการปรับปรุงการพยากรณ์จาก single exponential method โดยใช้ค่า weight ที่เป็นค่าคงที่สองค่าคือค่า α และค่า β โดยเป็นค่าคงที่ระหว่าง 0 กับ 1 ตามสมการดังนี้

$$L_t = \alpha Y_1 + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (2.6)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (2.7)$$

$$F_{t+m} = L_t + b_t m \quad (2.9)$$

โดยการพยากรณ์ที่คาบเวลาใด ๆ จะอยู่ในรูปของสมการเส้นตรงที่มีตัวแปร b_t แสดงถึงระดับความชันของเส้นตรงของสมการนี้

ตัวอย่าง:

แสดงวิธีการพยากรณ์ยอดขายโดยใช้วิธี Holt's Linear method ให้พยากรณ์ยอดขายตั้งแต่คาบเวลาที่ 25 ถึง 30 โดยสมมติให้ค่า $\alpha = 0.501$ และ $\beta = 0.072$

ตารางที่ 2.3 การพยากรณ์ยอดขายโดยใช้วิธี Holt's Linear method

	Period	Observed data	Smoothing of data	Smoothing of trend	Forecast when m=1	
	(t)	(Y_t)	(L_t)	(b_t)	(F_t)	
	1	143	143.00	9.00	-	
	2	152	152.00	9.00	152.00	
	3	161	161.00	9.00	161.00	
	4	139	154.47	7.88	170.00	
	5	137	149.65	6.97	162.35	
	6	174	165.33	7.59	156.62	
	7	142	157.43	6.48	172.92	
	8	141	152.43	5.65	163.91	
	9	162	160.05	5.79	158.08	
Test Set	10	180	172.93	6.30	165.84	
	11	164	171.60	5.75	179.24	
	12	171	174.17	5.53	177.36	
	13	206	192.88	6.47	179.70	
	14	193	196.17	6.25	199.35	
	15	207	204.71	6.41	202.41	
	16	218	214.57	6.66	211.12	
	17	229	225.12	6.94	221.23	
	18	225	228.52	6.68	232.06	
	19	204	219.57	5.56	235.21	
	20	227	226.07	5.63	225.13	
	21	223	227.34	5.31	231.69	
	22	242	237.33	5.65	232.65	
	23	239	240.99	5.51	242.98	
	24	266	256.27	6.21	246.49	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.3 (ต่อ)

	25				262.48	m=1
	26				268.69	m=2
	27				274.9	m=3
	28				274.9	m=4
	29				281.11	m=5
	30				287.32	m=6

จากตัวอย่าง กำหนดให้ค่าเริ่มต้น $L_1 = Y_1$ และ $b_1 = Y_2 - Y_1$ จะได้ค่าเริ่มต้นของ $L_1 = 143$ และ $b_1 = 9$

2.3 Regression method

Regression method เป็นวิธีการพยากรณ์โดยอาศัยความสัมพันธ์ของข้อมูลสองกลุ่ม เช่น ความสัมพันธ์ของยอดขายรถที่มีผลกับการใช้น้ำมัน โดยการทำ regression จะแบ่งเป็น simple regression โดยให้ทำนายค่าตัวแปร Y ที่การเปลี่ยนแปลงขึ้นอยู่กับค่าตัวแปร X หนึ่งตัวแปร และ multiple regression ที่การเปลี่ยนแปลงของค่าตัวแปร Y ขึ้นอยู่กับ ค่าตัวแปรในแนวแกน X หลายค่า (ค่าตัวแปร $X_1 ... X_k$)

2.3.1 Simple regression

เป็นการพยากรณ์ค่า Y โดยที่การเปลี่ยนแปลงของค่า X มีผลกับค่า Y โดยให้ความสัมพันธ์ของค่า X และ Y เป็นสมการเส้นตรงดังนี้

$$Y = a + bX + e \tag{2.10}$$

โดยค่าตัวแปร a และ b ต้องเป็นค่าที่ทำให้เกิดค่า sum square error (SSE) น้อยที่สุด ค่าของ error และ Sum square error หาได้จากสมการ

$$e_i = Y_i - \hat{Y}_i \tag{2.11}$$

โดยให้ Y_i เป็น ค่าจริงที่คาบเวลา i ใด ๆ

$\hat{Y}_i$ เป็นค่าที่พยากรณ์ได้ที่คาบเวลา i ใด ๆ

เมื่อได้ค่า error จะสามารถหาค่า sum square error(SSE) ได้จากสมการ

$$SSE = \sum_{i=1}^{n} e_i^2 \tag{2.12}$$

ค่า b ที่เป็น slop ของสมการเส้นตรงใน simple regression หาได้จากสมการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$b = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (2.13)$$

ค่า a ที่เป็นจุดตัดแกน Y ของสมการเส้นตรงใน simple regression หาได้จากสมการ

$$a = \bar{Y} - b\bar{X} \qquad (2.14)$$

ตัวอย่าง :

จากตารางที่ 2.4 จะแสดงข้อมูลระหว่าง ราคาสินค้า (P_i) และยอดการนำเข้าสินค้า (S_i) ดังนี้

ตารางที่ 2.4 ยอดการส่งออกสินค้าและราคาสินค้า

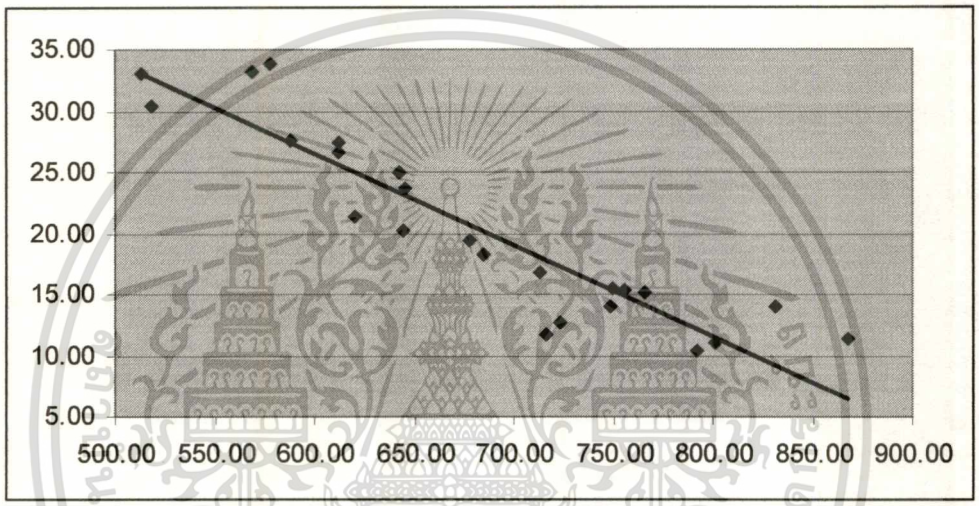
Shipments	Price
S_i	P_i
10.44	792.32
11.40	868.00
11.08	801.09
11.70	715.87
12.74	723.36
14.01	748.32
15.11	765.37
15.26	755.32
15.55	749.41
16.81	713.54
18.21	685.18
19.42	677.31
20.18	644.59
21.40	619.71
23.63	645.83
24.96	641.95
26.58	611.97
27.57	587.82
30.38	518.01
33.07	513.24
33.81	577.41
33.19	569.17
35.15	516.75
27.45	612.18
13.96	831.04

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากข้อมูลที่ได้ หาค่า a และ b ที่ทำให้ได้สมการเส้นตรงที่มีค่า SSE น้อยที่สุดดังนี้

$$S=71.7 - 0.075P$$

จากสมการ จะได้กราฟจากสมการเส้นตรงที่แสดงถึงเส้น trend line จากข้อมูลระหว่างราคาสินค้ากับจำนวนการนำเข้าสินค้าดังรูปที่ 2.1



รูปที่ 2.1 กราฟสมการเส้นตรงจากข้อมูลระหว่างราคาสินค้า กับจำนวนการนำเข้าสินค้า

2.3.2 The correlation coefficient

Correlation coefficient เป็นการวัดค่าความสัมพันธ์ระหว่างข้อมูลสองค่าที่ถูกนำมาใช้ในการพยากรณ์ด้วยวิธีการ regression ว่ามีความเกี่ยวข้องกันมากหรือน้อยอย่างไร โดยให้ข้อมูล 2 กลุ่มสมมติให้เป็นค่า X และ Y เราจะหาค่าของ correlation coefficient ระหว่าง 2 ค่านี้ จะได้ค่าของตัวแปร r_{xy} ที่เป็นค่าระหว่าง 1 และ -1 โดยค่าบวกและลบแสดงถึงความชันของเส้นตรงที่เป็นบวกหรือลบ และค่า $|r_{xy}|$ ที่ค่าเข้าใกล้ 1 มากเท่าไหร่ จะแสดงให้เห็นว่าค่า X และ Y มีความเกี่ยวข้องกันมาก ซึ่งตรงข้ามกับค่าที่เข้าใกล้ 0 แสดงให้เห็นว่าค่า X และ Y มีความเกี่ยวข้องกันน้อย ซึ่งหาได้จากสมการ

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (2.15)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{2.16}$$

$$Cov_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{2.17}$$

$$S_x^2 = Cov_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{2.18}$$

$$S_y^2 = Cov_{yy} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{2.19}$$

2.3.3 Coefficient of determination

Coefficient of determination เป็นการหาค่าของความสัมพันธ์ระหว่างค่าของข้อมูลจริง Y และค่าของข้อมูลที่พยากรณ์ได้ Y_i ที่มาจากการทำ regression โดย coefficient of determination จะอยู่ในรูปแบบของค่า R โดยค่า R^2 หาได้จากสมการ

$$R^2 = r_{xy}^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \tag{2.20}$$

โดยจะเขียนได้อีกรูปแบบหนึ่งที่เป็นค่าอัตราส่วนระหว่าง ค่า sum square regression และ sum square regression ดังนี้

$$\underbrace{\sum(Y_i - \bar{Y})^2}_{\text{Sum square total}} = \underbrace{\sum(Y_i - \hat{Y}_i)^2}_{\text{Sum square error}} + \underbrace{\sum(\hat{Y}_i - \bar{Y})^2}_{\text{Sum square regression}} \tag{2.21}$$

$$\text{SST} \qquad \text{SSE} \qquad \text{SSR}$$

$$R^2 = \frac{SSR}{SST}$$

ตัวอย่าง:

จากตารางที่ 2.4 จะแสดงการคำนวณค่า R^2 ระหว่าง ราคาสินค้า (P_i) และยอดการส่งออกสินค้า (S_i) ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.5 การคำนวณค่า R^2 ของราคาสินค้า (P_i) และยอดการส่งออกสินค้า (S_i)

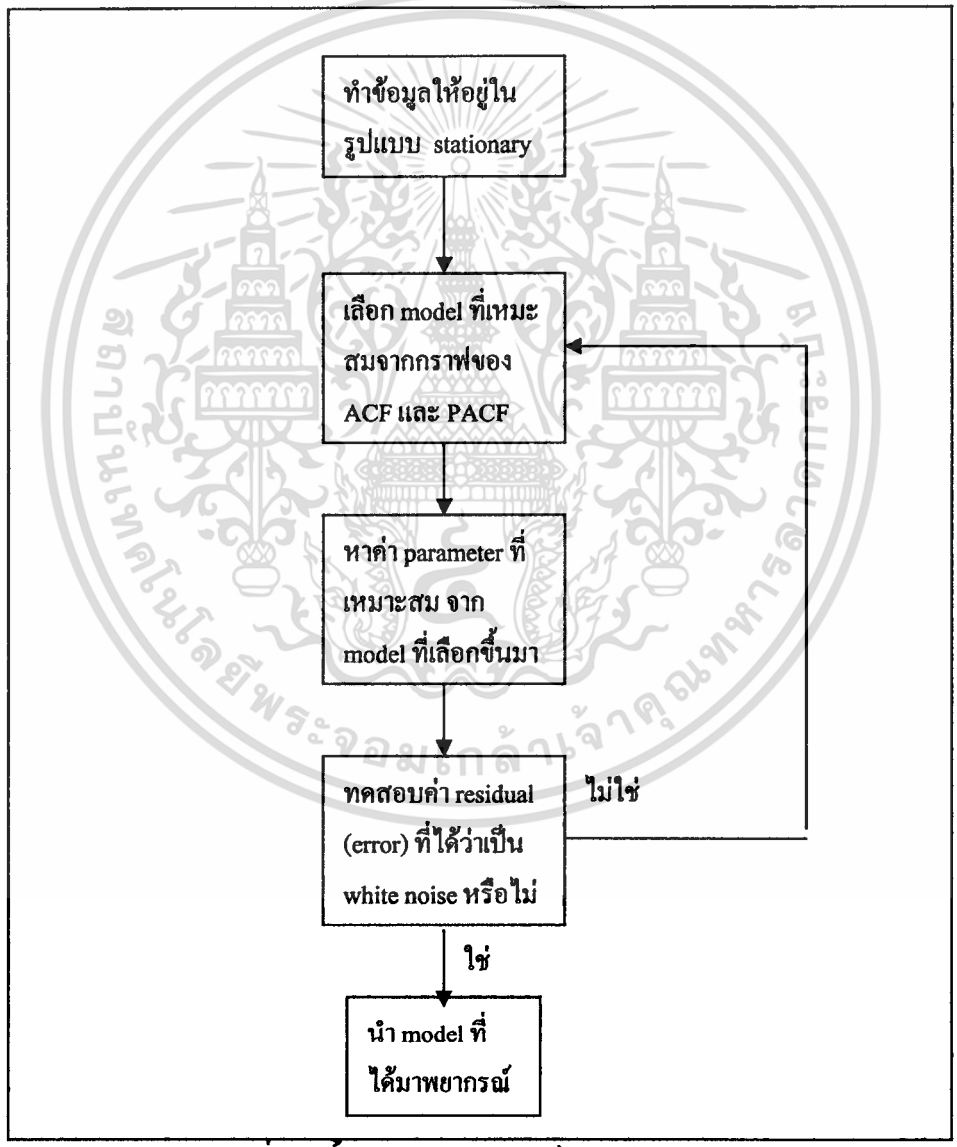
Price	Shipments	Estimate Shipment	Total deviation	Unexplained deviation	Explained deviation
X_i	Y_i	$\hat{Y}_i$	$Y_i - \overline{Y}$	$Y_i - \hat{Y}$	$\hat{Y}_i - \overline{Y}$
513.24	33.07	33.21	12.15	-0.14	12.29
516.75	35.15	32.94	14.23	2.21	12.02
518.01	30.38	32.85	9.46	-2.47	11.93
569.17	33.19	29.01	12.27	4.17	8.09
577.41	33.81	28.39	12.89	5.42	7.47
587.82	27.57	27.61	6.65	-0.04	6.69
611.97	26.58	25.80	5.66	0.78	4.88
612.18	27.45	25.79	6.53	1.66	4.87
619.71	21.40	25.22	0.48	-3.82	4.30
641.95	24.96	23.55	4.04	1.41	2.63
644.59	20.18	23.36	-0.74	-3.17	2.44
645.83	23.63	23.26	2.71	0.37	2.34
677.31	19.42	20.90	-1.50	-1.48	-0.02
685.18	18.21	20.31	-2.71	-2.10	-0.61
713.54	16.81	18.18	-4.11	-1.37	-2.74
715.87	11.70	18.01	-9.22	-6.31	-2.91
723.36	12.74	17.45	-8.18	-4.70	-3.47
748.32	14.01	15.58	-6.91	-1.56	-5.34
749.41	15.55	15.49	-5.37	0.06	-5.43
755.32	15.26	15.05	-5.66	0.21	-5.87
765.37	15.11	14.30	-5.81	0.82	-6.62
792.32	10.44	12.28	-10.48	-1.84	-8.64
801.09	11.08	11.62	-9.84	-0.54	-9.30
831.04	13.96	9.37	-6.96	4.59	-11.55
868.00	11.40	6.60	-9.52	4.80	-14.32
	20.92		1566.16	209.35	1352.73
	$\overline{Y}$		SST	SSE	SSR

จากตารางเราจะได้ค่า $R^2 = \frac{1352.73}{1566.2} = 86.3\%$ ซึ่งแสดงให้เห็นว่า ค่าของ Y มีความสัมพันธ์เกี่ยวข้องกับค่า X มาก ซึ่งทำให้สมการเส้นตรงที่ได้จากการทำ regression สามารถอธิบายความสัมพันธ์ระหว่างราคาสินค้า และยอดการนำเข้าสินค้าได้ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 ARIMA model

การพยากรณ์โดยใช้วิธีการของ Box-Jenkins หรือ ARIMA model เป็นการนำชุดข้อมูลที่จะพยากรณ์มาแปลงให้มีความเป็น stationary แล้วจึงนำมาพิจารณาจากค่าของ auto correlation function (ACF) และค่า partial autocorrelation coefficient (PACF) ว่ารูปแบบที่ได้ มีความเหมาะสมกับการใช้ autoregressive model , moving average model หรือ ใช้ผสมกันทั้งสอง model โดยมีขั้นตอนดังรูปที่ 2.2



รูปที่ 2.2 ขั้นตอนการพยากรณ์โดยใช้ ARIMA model

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.1 The autocorrelation function (ACF)

Autocorrelation function เป็นการหาค่าของความสัมพันธ์ในตัวมันเองของข้อมูลในช่วงห่างต่าง ๆ กัน $(r_1, r_2 .....)$ เช่น r_2 แสดงค่าความสัมพันธ์ระหว่างข้อมูลที่ช่วงเวลา t และ t-2 โดยหาได้จากสมการ

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \qquad (2.22)$$

จากสมการที่ได้ ชุดของข้อมูลที่ไม่ค่อยมีความสัมพันธ์กันจะมีค่าเข้าใกล้ 0 ค่าของ ACF จะถูกพิจารณาจากความน่าจะเป็น 95% ที่จะเป็นค่า white noise โดยพิจารณาจากค่า $\frac{\pm 1.96}{\sqrt{n}}$ โดยให้ n เป็นจำนวนของข้อมูล

2.4.2 The partial autocorrelation function (PACF)

partial autocorrelation function จะนำมาใช้ในการแสดงค่าของ degree ระหว่างข้อมูล Y_t และ y_{t-k} โดยหาค่าได้จากการทำ multiple regression ภายในตัวมันเองของค่า Y_t กับค่า $Y_{t-1}, ..., Y_{t-k}$ จากสมการ

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + ... + b_k Y_{t-k} \qquad (2.23)$$

ค่า PACF α_k หาได้จาก coefficient b_k จากสมการ multiple regreaaion

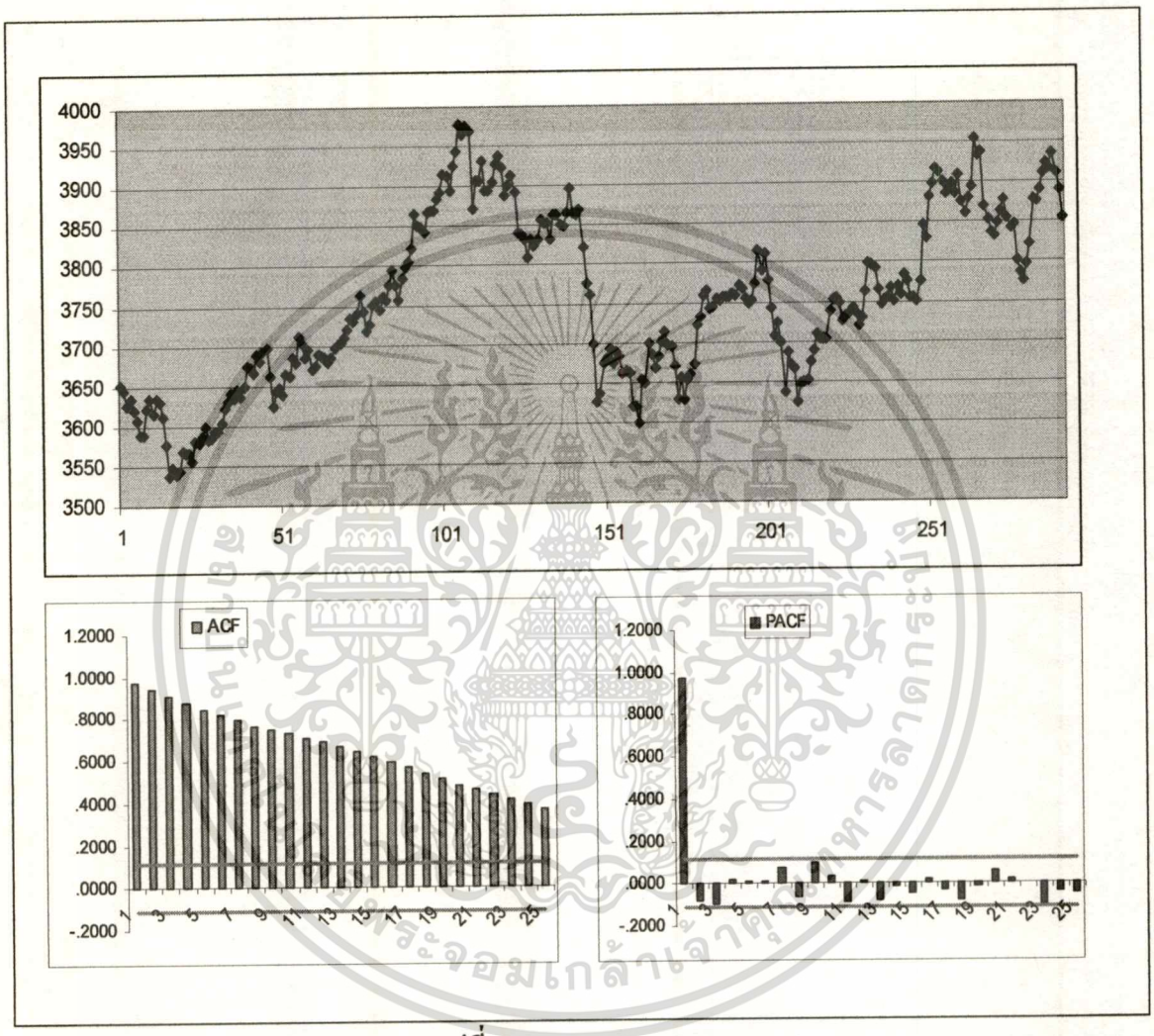
2.4.3 การทำข้อมูลให้เป็น stationary

ข้อมูลที่มีความเป็น stationary จะเป็นข้อมูลที่มีการกวัดแกว่งอยู่ที่ค่าคงที่ค่าหนึ่ง โดยจะไม่มีการกวัดแกว่งไปในระดับที่เพิ่มขึ้น หรือลดลง โดยค่า mean และ variance ในแต่ละช่วงของข้อมูลจะคงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง:

ข้อมูลของ Dow-Jones index จากรูปที่ 2.3 จะเห็นได้ว่า ไม่มีความเป็น stationary โดยจะมีการ กวัดแกว่งที่มีแนวโน้มเพิ่มขึ้น และลดลง ค่าของ ACF ที่ได้ จะมีค่าสูงมากที่ lag_1 และค่อย ๆ ลดลงเรื่อย ๆ ในส่วนของ PACF จะมีค่าสูงมากเข้าใกล้ 1 ที่ lag_1



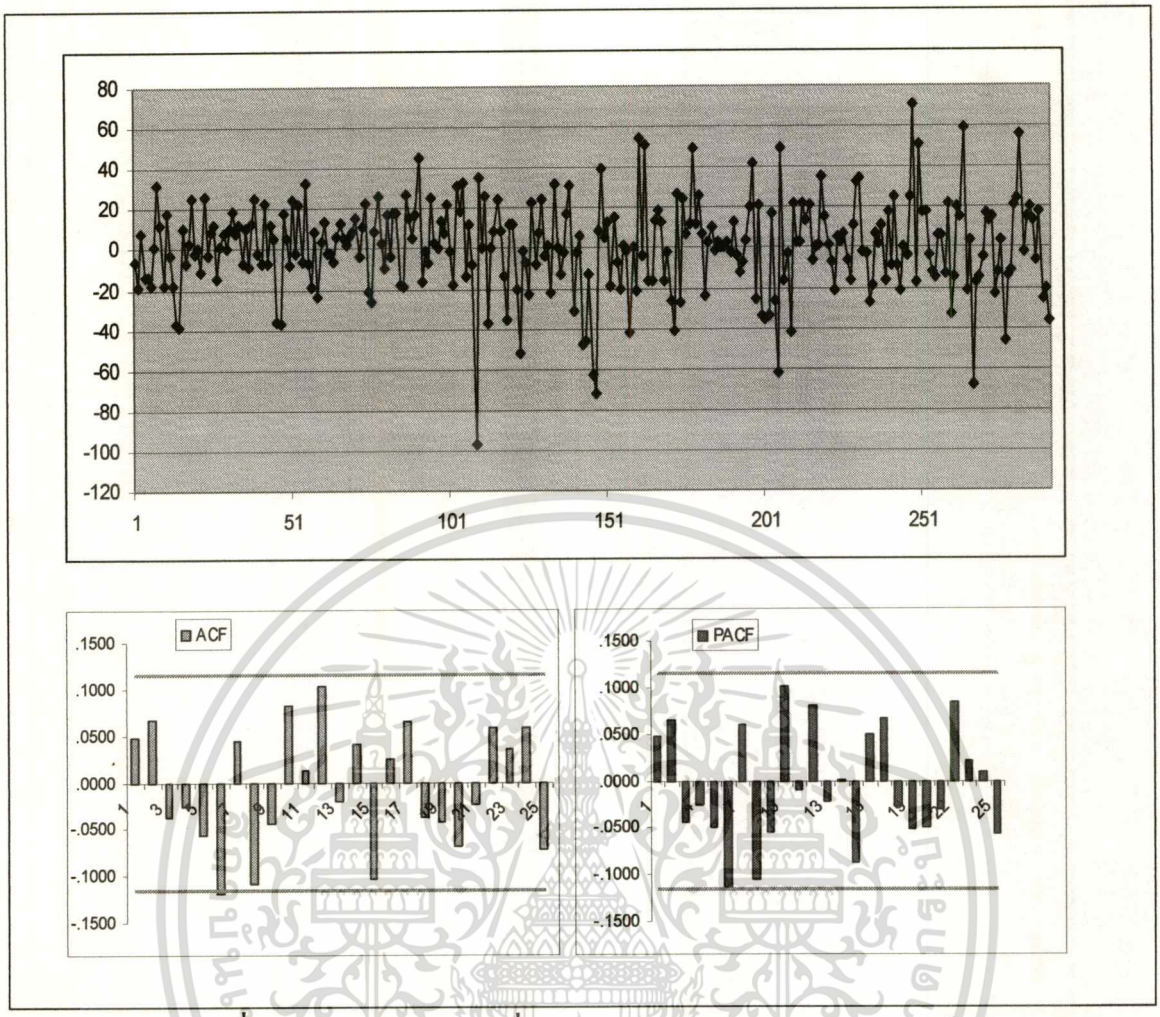
รูปที่ 2.3 Dow-Jones index

เราสามารถกำจัดข้อมูลที่เป็น non-stationary ออก โดยใช้การ differencing ดังสมการ

$$Y' = Y_t - Y_{t-1} \tag{2.24}$$

ซึ่งวิธีการนี้เราจะเรียกว่า first-order differences จากข้อมูล Dow-Jones index เมื่อผ่านการทำ first-order differences จะได้ข้อมูลที่เป็น stationary ดังรูปที่2.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.4 Dow-Jones index ที่ผ่านการทำ first-order differencing แล้ว

ในกรณีที่ข้อมูลผ่านการทำ first-order differencing แล้วยังไม่เป็น stationary เราจะเอาข้อมูลที่ได้มาทำ differencing อีกครั้ง เรียกว่า second-order differences

$$Y'' = Y'_t - Y'_{t-1} \tag{2.25}$$

ในกรณีที่ข้อมูลมีความเป็น seasonal เราจะนำมาผ่านการทำ seasonal-differenes ก่อนดังนี้

$$Y_t' = Y_t - Y_{t-12} \tag{2.26}$$

จากข้อมูลที่ได้ ถ้ายังไม่มีความเป็น stationary เราจะนำมาทำ first-order differenes อีกครั้งเพื่อให้ข้อมูลเข้าใกล้ความเป็น stationaryดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Y_t^* = Y'_t - Y'_{t-1}$$
$$= (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13})$$

2.4.4 Backshift notation

เราสามารถนำวิธีการ backward shift มาใช้ในสมการที่มีการ differencing เพื่อให้ง่ายต่อการเขียนดังนี้

$$BY_t = Y_{t-1} \tag{2.27}$$
$$B(BY_t) = B^2 Y_t = Y_{t-2} \tag{2.28}$$

2.4.5 การนำ ARIMA model มาใช้กับข้อมูล time series

ARIMA model จะเขียนอยู่ในรูปแบบของ ARIMA(p,d,q) โดยค่า parameter แต่ละค่าจะมีความหมายดังนี้

AR: p = order ของการทำ autoregressive

I: d = degree ของการทำ first differencing

MA: q = order ของการทำ moving average

2.4.6 Autoregressive model ลำดับที่ 1

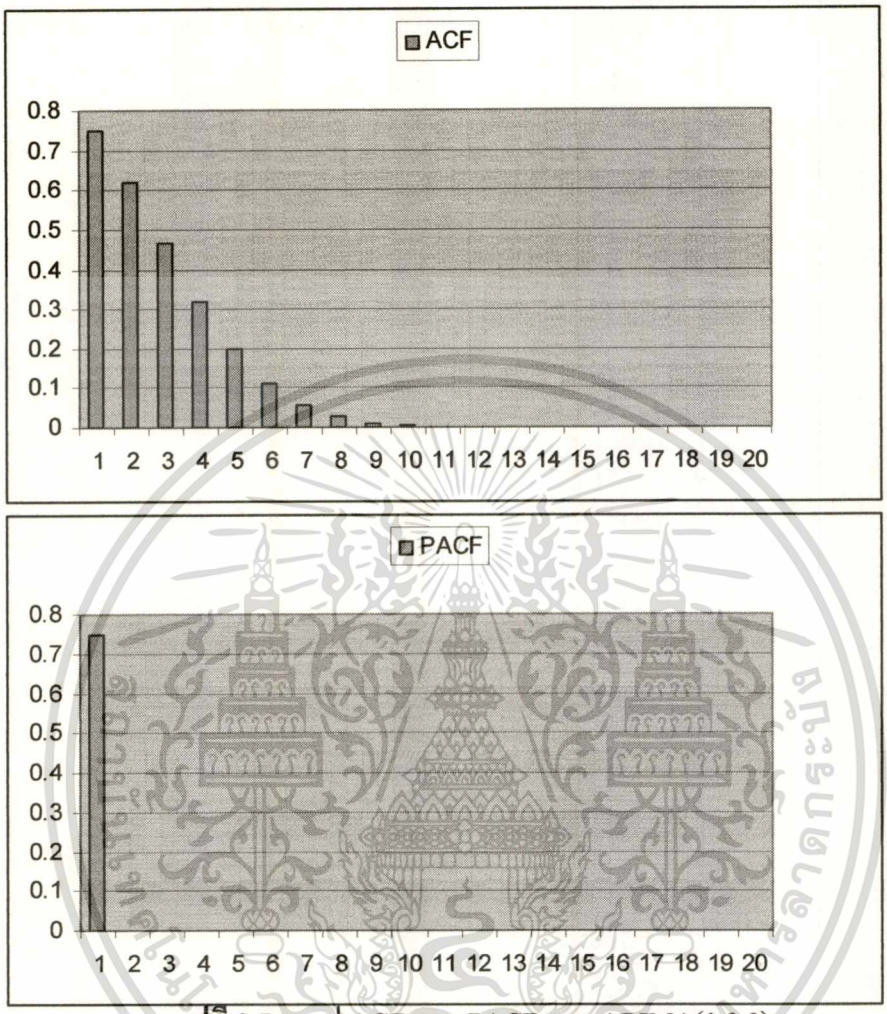
Autoregressive model ลำดับที่ 1 หรือ ARIMA(1,0,0) ค่าของข้อมูลที่ทำการพยากรณ์ Y_t จะขึ้นอยู่กับค่าในอดีต Y_{t-1} โดยมีสมการดังนี้

$$Y_t = C + \theta_1 Y_{t-1} + e_t \tag{2.29}$$

กราฟของ ACF และ PACF จะมีรูปแบบใกล้เคียงกับรูปที่ 2.5 โดยจากกราฟ ค่าของ ACF จะลดลงอย่างรวดเร็วในลักษณะของ exponential เข้าใกล้ 0 และค่าของ PACF ที่ lag_1 จะมีค่าสูงมาก เข้าใกล้ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง



รูปที่ 2.5 กราฟ ACF และ PACF ของ ARIMA(1,0,0)

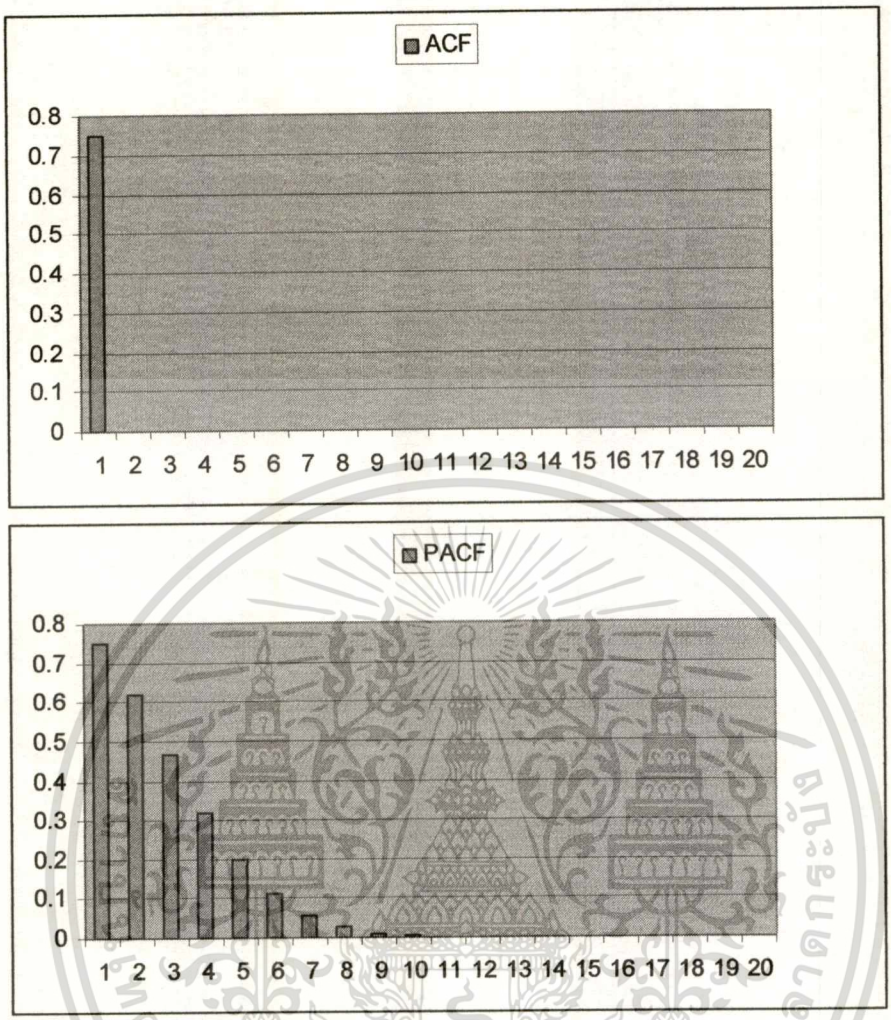
2.4.7 Moving average model ลำดับที่ 1

Moving average model ลำดับที่ 1 หรือ ARIMA(0,0,1) ค่าของข้อมูลที่พยากรณ์ได้ Y_t จะขึ้นอยู่กับค่า error e_t และค่า error ในอดีต e_{t-1} โดยเขียนเป็นสมการได้ดังนี้

$$Y_t = c + e_t - \theta_1 e_{t-1} \tag{2.30}$$

โดยกราฟของ ACF และ PACF จะมีรูปแบบใกล้เคียงกับรูปที่ 2.6 ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
04902



รูปที่ 2.6 กราฟ ACF และ PACF ของ ARIMA(0,0,1)

จากกราฟ จะแสดงให้เห็นว่า ค่าของ PACF จะลดลงเข้าใกล้ 0 แบบ exponential และค่าของ ACF จะมีค่าสูง เข้าใกล้ 1 ที่ lag_1

2.4.8 Autoregressive model ในลำดับที่สูงขึ้น

autoregressive ในลำดับที่ p ใด ๆ หรือ ARIMA(p,0,0) จะเขียนเป็นสมการ

$$Y_t = c + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + ... + \theta_p Y_{t-p} + e_t \qquad (2.31)$$

โดยค่าของ ACF จะลดลงเข้าใกล้ 0ในแบบ exponential หรืออยู่ในรูปแบบของ sine-wave และ PACF จะมีค่าสูงมากที่ lag_1 ถึง lag_p

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.9 Moving average model ในลำดับที่สูงขึ้น

Moving average model หรือ ARIMA(0,0,1) จะเขียนเป็นสมการได้ดังนี้

$$Y_t = c + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - .... - \theta_q e_{t-q} \tag{2.32}$$

โดยค่าของ PACF จะลดลงเข้าใกล้ 0 ในแบบ exponential หรืออยู่ในรูปแบบ sine-wave และค่า ACF จะมีค่าสูงมากที่ lag_1 ถึง lag_q

2.4.10 ARIMA model ในแบบผสม

ข้อมูลที่นำมาใช้พยากรณ์ บางครั้ง model ที่เหมาะสมกับการพยากรณ์ จะเป็น model ที่ผสมกันระหว่าง autoregressive model และ moving average model ในรูปแบบของ ARIMA(p,0,q) โดยจะเขียนให้อยู่ในรูปแบบของสมการ

$$Y_t = c + \theta_1 Y_{t-1} + ... + \theta_p Y_{t-p} + e_t - \phi_1 e_{t-1} - ... - \phi_q e_{t-q} \tag{2.33}$$

หรืออยู่ในรูปแบบ back shift notation ดังนี้

$$(1 - \theta_1 B - ... - \theta_p B^p) Y_t = c + (1 - \phi_1 B - ... - \phi_q B^q) e_t \tag{2.34}$$

ในกรณีที่ข้อมูลไม่มีความเป็น stationary เราจะนำวิธีการ differencing มาใช้ โดยจะอยู่ในรูปของ ARIMA(p,d,q) ตัวอย่างเช่น ARIMA(1,1,1) จะเขียนได้เป็นสมการในรูปแบบของ back shift notation ดังนี้

$$(1 - \theta_1 B)(1 - B) Y_t = c + (1 - \phi_1 B) e_t \tag{2.35}$$

ในกรณีที่ข้อมูลมีความเป็น seasonal จะเขียนให้อยู่ในรูปแบบของ $ARIMA(p,d,q)(P,D,Q)_s$ โดยให้ p,d,q แทน model ที่เป็น non-seasonal และ P,D,Q แทน model ที่เป็น seasonal

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง:

model ของ $ARIMA(1,1,1)(1,1,1)_4$ จะเขียนเป็นสมการได้ดังนี้

$$(1-\theta_1 B)(1-\Phi_1 B^4)(1-B)(1-B^4)Y_t = (1-\phi_1 B)(1-\Theta_1 B^4)e_t \qquad (2.36)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

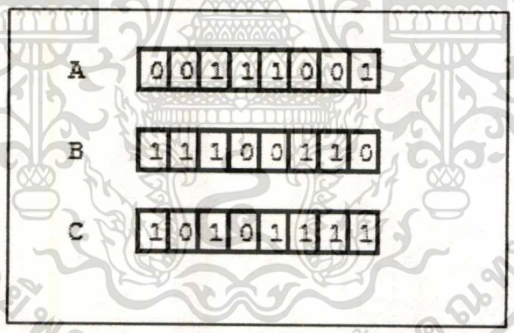
บทที่ 3

Genetic Algorithm

Genetic algorithm เป็นวิธีที่ใช้ในการคัดเลือกคำตอบที่ดีที่สุดของปัญหาที่เราสนใจ โดยมีวิธีการเลือกข้อมูลที่จะนำมาเป็นประชากร และมีการคัดเลือกเพื่อที่จะนำมาสร้างประชากรรุ่นถัดไป ซึ่งเป็นการจำลองมาจากรูปแบบของวิธีการทางพันธุศาสตร์ที่มีการคัดประชากรที่อ่อนแอออกแล้ว จึงได้เป็นประชากรที่มีลักษณะดีกว่าในรุ่นต่อไปนั่นเอง

3.1 หลักการของ Genetic algorithm

หลักการของ Genetic algorithm ในการใช้หาคำตอบ เราจะมีการกำหนดตัวแปรซึ่งใช้เป็นตัวแทนของคำตอบให้อยู่ในรูปของโครโมโซม โดยเก็บเป็นสายอักขระของตัวเลขดังรูปที่ 3.1 ที่แสดงตัวอย่างของโครโมโซมในรูปแบบของเลขฐานสอง



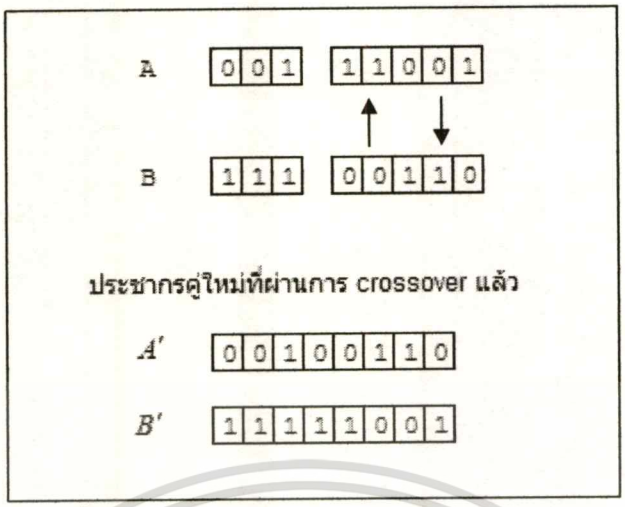
รูปที่ 3.1 ตัวอย่างของโครโมโซมในรูปแบบของเลขฐานสอง

การทำจะทำให้โครโมโซมเกิด ลักษณะใหม่ที่ดีนั้น จะผ่านขั้นตอนของการ crossover และ การ mutation ซึ่งจะทำให้เกิดเป็นโครโมโซมลักษณะใหม่ที่ต่างจากประชากรรุ่นพ่อแม่ดังนี้

- กระบวนการ Crossover

กระบวนการ crossover เป็นการแลกเปลี่ยนคุณลักษณะของประชากรสองตัว โดยจะเป็นการไขว้กันของโครโมโซมตามตำแหน่งที่เรากำหนด ดังรูปที่ 3.2

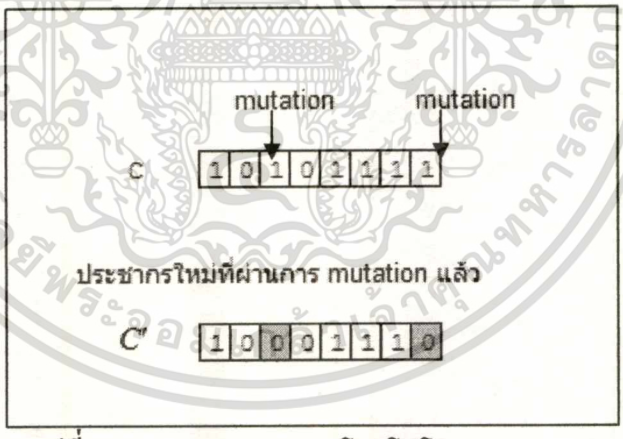
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 การ crossover ของโครโมโซม A และ B

- กระบวนการ Mutation

กระบวนการ mutation เป็นการเปลี่ยนแปลงลักษณะบางประการภายในตัวโครโมโซมเอง ให้ได้ประชากรใหม่ที่มีลักษณะต่างจากต้นแบบดังรูปที่3.3



รูปที่ 3.3 การ mutation ของโครโมโซม C

หลังจากผ่านขั้นตอนการ crossover และ mutation แล้ว เราจะนำค่าที่ได้ของโครโมโซมมาบวกเข้าด้วยกัน ได้เป็นค่า fitness value และนำโครโมโซมแต่ละตัวมาหาค่า probability โครโมโซมที่มีค่า probability สูง ก็มีความเป็นไปได้สูงที่จะถูกคัดเลือกให้เป็นประชากรรุ่นต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ขั้นตอนของ genetic algorithm

ขั้นตอนของ genetic algorithm นั่น จะเริ่มต้นด้วยการสุ่มเพื่อหาประชากรต้นแบบ แล้วทำการคัดเลือกประชากรรุ่นแรกจากค่า probability ของแต่ละโครโมโซมโดยคำนวณจากค่าของ fitness value โครโมโซมที่มีค่า probability สูง ก็จะมีความเป็นไปได้ที่จะถูกเลือกให้เป็นประชากรรุ่นแรกด้วย

หลังจากได้ประชากรรุ่นที่หนึ่งออกมาแล้ว ก็นำมาผ่านขั้นตอนของการ crossover และ mutation แล้วนำมาคัดเลือกเพื่อเป็นประชากรรุ่นที่สองโดยหาค่าของ fitness และ probability ในแต่ละขั้นตอน จะยกตัวอย่างการนำ generic algorithm มาแก้ไขปัญหา โดยใช้หาค่าสูงสุดของ function $f(x) = 15x - x^2$ ดังนี้

3.2.1 การคัดเลือกประชากรต้นแบบ

ประชากรที่เรานำมาใช้คือค่า X ซึ่งจะมีค่าตั้งแต่ 0 ถึง 15 โดยแปลงให้อยู่ในรูปของเลขฐานสองดังนี้

ค่า X	โครโมโซม	ค่า X	โครโมโซม	ค่า X	โครโมโซม	ค่า X	โครโมโซม
0	0000	4	0100	8	1000	12	1100
1	0001	5	0101	9	1001	13	1101
2	0010	6	0110	10	1010	14	1110
3	0011	7	0111	11	1011	15	1111

ซึ่งจะได้ประชากรต้นแบบ v_i ออกมาดังนี้

$$v_0 = (0000)$$
$$v_1 = (0001)$$
$$v_2 = (0010)$$
$$v_3 = (0011)$$
$$v_4 = (0100)$$
$$v_5 = (0101)$$
$$v_6 = (0110)$$
$$v_7 = (0111)$$
$$v_8 = (1000)$$
$$v_9 = (1001)$$
$$v_{10} = (1010)$$
$$v_{11} = (1011)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$v_{12} = (1100)$$
$$v_{13} = (1101)$$
$$v_{14} = (1110)$$
$$v_{15} = (1111)$$

นำค่า v_i หรือค่า X ที่ได้ไปแทนค่า ใน function $f(x) = 15x - x^2$ จะได้ค่าของผลลัพธ์ $eval_{(v_i)}$ ซึ่งเรานำมาใช้หาค่า fitness (F) ตามสมการ

$$F = \sum_{i=0}^{15} eval_{(v_i)} \tag{3.1}$$

จะได้ค่า fitness F = 560 นำค่าที่ได้มาหาค่าของ probability จากสมการ

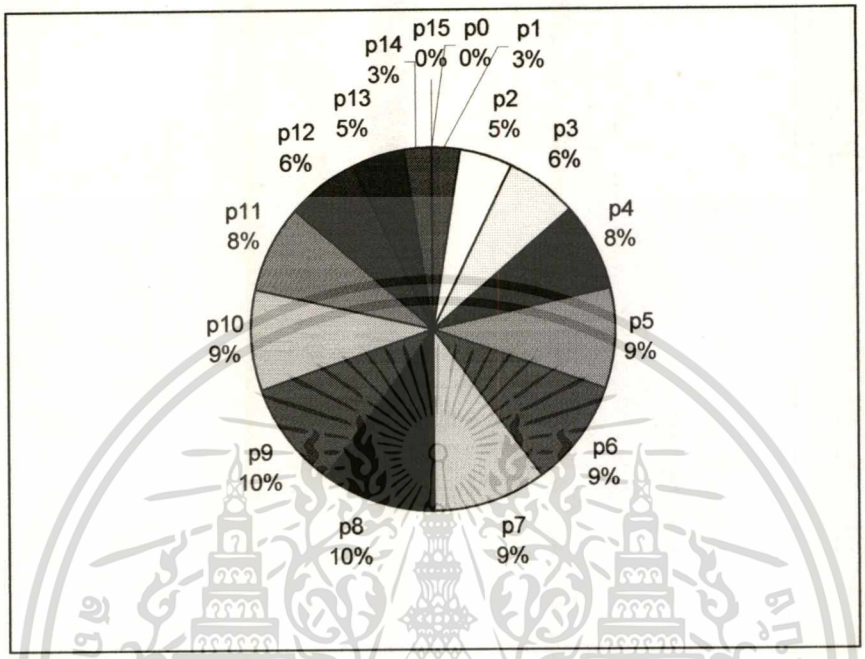
$$p_i = eval(v_i) / F \tag{3.2}$$

จะได้ค่า probability ของแต่ละโครโมโซม ดังนี้

$$p_0 = 0.000000$$
$$p_1 = 0.025000$$
$$p_2 = 0.046429$$
$$p_3 = 0.064286$$
$$p_4 = 0.078571$$
$$p_5 = 0.089286$$
$$p_6 = 0.096429$$
$$p_7 = 0.100000$$
$$p_8 = 0.100000$$
$$p_9 = 0.096429$$
$$p_{10} = 0.089286$$
$$p_{11} = 0.078571$$
$$p_{12} = 0.064286$$
$$p_{13} = 0.046429$$
$$p_{14} = 0.025000$$
$$p_{15} = 0.000000$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากค่า probability ที่ได้ จะได้ graph ของ roulette wheel selection ดังรูปที่ 3.3



รูปที่ 3.4 Roulette wheel selection

นำค่าที่ได้ไปแปลงเป็นค่า cumulative q_i ดังนี้

q0	0.000000
q1	0.025000
q2	0.071429
q3	0.135714
q4	0.214286
q5	0.303571
q6	0.400000
q7	0.500000
q8	0.600000
q9	0.696429
q10	0.785714
q11	0.864286
q12	0.928571
q13	0.975000
q14	1.000000
q15	1.000000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้น เราจะหาค่า random 16 ค่า ที่อยู่ในช่วง [0..1] ดังนี้

0.404340	0.671941	0.072595	0.722103
0.642070	0.081619	0.116802	0.657725
0.427421	0.683723	0.389427	0.376860
0.882446	0.106756	0.532196	0.096038

นำค่า random ที่ได้ มาเปรียบเทียบกับค่า cumulative q_i แต่ละตัว เพื่อคัดเลือกไปเป็นประชากรรุ่นที่ 1 เช่น ค่าแรก 0.404340 มีค่าอยู่ระหว่าง q6 และ q7 ดังนั้น โครโมโซม q7 จะถูกคัดเลือกเข้ามาเป็นประชากรรุ่นที่หนึ่ง ทำตามขั้นตอนนี้ไปจบครบ 16 ค่า เราจะได้ประชากรรุ่นที่หนึ่ง 16 ค่า ดังนี้

v7 -> v'1	(0111)
v9 -> v'2	(1001)
v3 -> v'3	(0011)
v10 -> v'4	(1010)
v9 -> v'5	(1001)
v3 -> v'6	(0011)
v3 -> v'7	(0011)
v9 -> v'8	(1001)
v7 -> v'9	(0111)
v9 -> v'10	(1001)
v6 -> v'11	(0110)
v6 -> v'12	(0110)
v12 -> v'13	(1100)
v3 -> v'14	(0011)
v8 -> v'15	(1000)
v3 -> v'16	(0011)

เมื่อได้ประชากรรุ่นที่หนึ่งออกมา เราจะนำมาผ่านกระบวนการ crossover และ mutation ในหัวข้อถัดไป

3.2.2 กระบวนการ crossover

ก่อนจะมีการ crossover เราต้องมีการกำหนดเปอร์เซ็นต์ในการเลือกโครโมโซมที่จะนำมา crossover ซึ่งในที่นี้ เราจะเลือก crossver 25% หมายถึง โครโมโซมจำนวน 4 ตัวจะถูกเลือกแบบสุ่มมาทำ crossover

สมมติว่า เราสุ่มได้โครโมโซม 4 ตัว ดังนี้ คือ v'4 , v'6 , v'10 , v'15 นำมาจับคู่เพื่อทำ cross over โดยให้คู่ v'4 และ v'6 crossover ที่ตำแหน่งที่สอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

v'4	(10|10)
v'6	(00|11)

จะได้ค่าดังนี้

v''4	(1011)
v''6	(0010)

และให้คู่ v'10 และ v'15 crossover ที่ตำแหน่งที่หนึ่ง

v'10	(1|001)
v'15	(1|000)

จะได้ค่าดังนี้

v''10	(11000)
v'15	(1001)

หลังจากที่เราทำการ crossover เราจำทำขั้นตอนของการ mutation ต่อไป

3.2.3 กระบวนการ mutation

กระบวนการ mutation นั้น เราต้องมีการกำหนดเปอร์เซ็นต์ของการ mutation โดยส่วนใหญ่จะกำหนดค่าให้น้อย ๆ โดยจากตัวอย่าง เราจะกำหนอดให้มีการ mutation 1% ของข้อมูลทั้งหมด โดยวัดจากข้อมูลเป็นจำนวน bit ดังนี้

จำนวน bit = 16*4 = 64

จำนวน bit ที่จะทำ mutation = 64*0.01 = 0.64 หรือ 1 bit นั่นเอง

จากนั้น เราจะสุ่มเลือกมา 1 โครโมโซม เพื่อที่จะทำการ mutation 1 bit สมมติให้เป็นโครโมโซม v'5 ตำแหน่งที่สอง

v'5	(1001)

จะได้ค่า v''5 ที่ผ่านการ mutation ดังนี้

v''5	(1101)

หลังจากที่เราทำการ crossover และ mutation แล้ว เราจะนำข้อมูลที่ได้ไปทำหาค่าของ probability เพื่อที่จะคัดเลือกเป็นประชากรรุ่นที่สองต่อไป โดยทำขั้นตอนซ้ำจนกว่าจะได้ค่าที่ดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การประยุกต์ใช้งาน Autoregressive model ร่วมกับ Genetic Algorithm

เราสามารถนำวิธีการ Genetic algorithm มาช่วยในการหา Autoregressive model ของการ forecasting ได้ โดย genetic algorithm จะทำหน้าที่ในการคัดเลือก Autoregressive model ที่ดีที่สุดในการทำ forecasting โดยคัดจาก model ที่เริ่มต้นจากการสุ่ม และเลือกเอากลุ่มของ Autoregressive model ที่มีความคลาดเคลื่อนน้อยมาปรับปรุงเป็นประชากรในรุ่นถัดไป และทำซ้ำจนกว่าจะได้ Autoregressive model ที่ดีที่สุดในการใช้งาน

4.1 การแทนค่าโครโมโซมของ Autoregressive model ใน Generic algorithm

หลักการทำงานของ Genetic algorithm เพื่อใช้ในการหา Autoregressive model นั้น เราจะใช้ Autoregressive model ในแต่ละ model เป็นโครโมโซมของ Genetic algorithm ดังตัวอย่างโครโมโซม (10011) ซึ่งเป็นเลขฐาน 2 จำนวน 5 bit เพื่อแทนค่า $AR(p)$ โดยจากโครโมโซมตัวอย่าง จะได้ Autoregressive model เป็น $AR(1,4,5)$

4.2 การทำข้อมูลให้เป็น stationary

ก่อนที่เราจะนำข้อมูลไปเข้า Genetic algorithm นั้น ต้องมีการทำข้อมูลให้เป็น stationary ก่อน เราสามารถกำจัดข้อมูลที่เป็น non-stationary ออก โดยใช้การ differencing ดังสมการ

$$Y' = Y_t - Y_{t-1} \tag{4.1}$$

ซึ่งวิธีการนี้เราจะเรียกว่า first-order differences ในกรณีที่ข้อมูลผ่านการทำ first-order differencing แล้วยังไม่เป็น stationary เราจะเอาข้อมูลที่ได้มาทำ differencing อีกครั้ง เรียกว่า second-order differences ดังสมการ

$$Y'' = Y'_t - Y'_{t-1} \tag{4.2}$$

ซึ่ง order ของการทำ differencing นั้น เราสามารถเพิ่มได้เรื่อย ๆ จนกว่าข้อมูลจะมีความเป็น stationary

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การทำงานของ Generic algorithm

4.1 สุ่มค่าประชากรเริ่มต้น

ในส่วนของโปรแกรมจะสุ่มค่าของโครโมโซมที่เป็นประชากรเริ่มต้นมา 8 ค่าดังตัวอย่าง

V0	(11010)	AR(1,2,4)
V1	(01000)	AR(2)
V2	(11000)	AR(1,2)
V3	(01100)	AR(2,3)
V4	(10110)	AR(1,3,4)
V5	(11110)	AR(1,2,3,4)
V6	(10110)	AR(1,3,4)
V7	(11110)	AR(1,2,3,4)

4.2 แทนค่าโครโมโซม

จากโครโมโซมที่ได้ จะนำมาแปลงเป็นสมการ autoregressive และหาผลลัพธ์ที่ได้จากสมการ คือค่า 1/SSE โดย SSE (Sum Square Error) ได้จากการนำแต่ละสมการมาทำการ forecast แล้วได้ค่า error ออกมา

Eval0	0.002458746173733
Eval1	0.002472255665248
Eval2	0.002472256189243
Eval3	0.002456781047412
Eval4	0.002458746173733
Eval5	0.002458756176736
Eval6	0.002458746173733
Eval7	0.002458756176736

4.3 หาค่า fitness และค่า ค่า probability ของแต่ละประชากร

ค่า fitness จากสมการ $F = \sum_{i=0}^{7} eval_{(v_i)}$ ซึ่งเป็นผลรวมของผลลัพธ์ที่ได้จากการทำนาย ได้ค่า fitness F=0.019695043776574

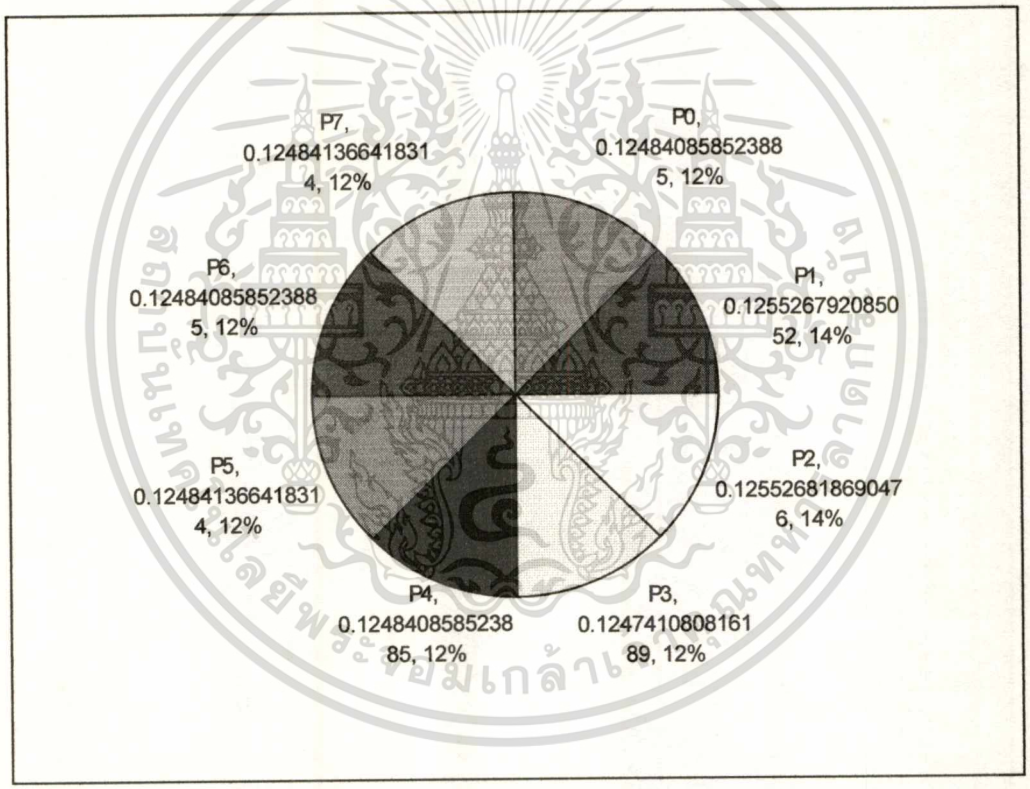
หาค่า probability จากค่า fitness ที่ได้โดยใช้สมการ $p_i = eval(v_i)/F$ จะได้ค่ probability ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

P0	0.124840858523885
P1	0.125526792085052
P2	0.125526818690476
P3	0.124741080816189
P4	0.124840858523885
P5	0.124841366418314
P6	0.124840858523885
P7	0.124841366418314

4.4 สร้าง roulette wheel selection และหาค่า cumulative เพื่อใช้ในการคัดเลือกประชากร

นำค่า probability ที่ได้ มาสร้าง roulette selection และค่า cumulative ได้ดังนี้



รูปที่ 4.1 Roulette wheel selection ที่คำนวณจากค่า probability

Q0	0.124840858523885
Q1	0.250367650608937
Q2	0.375894469299414
Q3	0.500635550115603
Q4	0.625476408639488
Q5	0.750317775057801
Q6	0.875158633581687
Q7	1.000000000000000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 สุ่มคัดเลือกประชากร

จากนั้น เราจะหาค่า random 8 ค่า ที่อยู่ในช่วง [0..1] ดังนี้

0.404340	0.671941	0.072595	0.722103
0.642070	0.081619	0.116802	0.657725

นำค่า random ที่ได้ มาเปรียบเทียบกับค่า cumulative Qi แต่ละตัว เพื่อคัดเลือกไปเป็นประชากรรุ่นที่ 1 เช่น ค่าแรก 0.404340 มีค่าอยู่ระหว่าง Q2 และ Q3 ดังนั้น โครโมโซม Q3 จะถูกคัดเลือกเข้ามาเป็นประชากรรุ่นที่หนึ่ง ทำตามขั้นตอนนี้ไปจบครบ 8 ค่า เราจะได้ประชากรรุ่นที่หนึ่ง 8 ค่า ดังนี้

V3 -> V'0	(01100)
V5 -> V'1	(11110)
V0 -> V'2	(11010)
V5 -> V'3	(11110)
V5 -> V'4	(11110)
V0 -> V'5	(11010)
V0 -> V'6	(11010)
v5 -> V'7	(11110)

เมื่อได้ประชากรรุ่นที่หนึ่งออกมา เราจะนำมาผ่านกระบวนการ crossover และ mutation ในหัวข้อถัดไป

3.2.4 กระบวนการ crossover

ก่อนจะมีการ crossover เราต้องมีการกำหนดเปอร์เซ็นต์ในการเลือกโครโมโซมที่จะนำมา crossover ซึ่งในที่นี้ เราจะเลือก crossver 25% หมายถึง โครโมโซมจำนวน 2 ตัวจะถูกเลือกแบบสุ่มมาทำ crossover

สมมติว่า เราสุ่มได้โครโมโซม 2 ตัว ดังนี้ คือ V'4 , V'6 นำมาจับคู่เพื่อทำ cross over โดยให้คู่ V'0 และ V'1 crossover ที่ตำแหน่งที่สอง

V'0	(01|100)
V'1	(11|110)

จะได้ค่าดังนี้

v''0	(01110)
v''1	(11100)

หลังจากที่เราทำการ crossover เราจำทำขั้นตอนของการ mutation ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.5 กระบวนการ mutation

กระบวนการ mutation นั้น เราต้องมีการกำหนดเปอร์เซ็นต์ของการ mutation โดยส่วนใหญ่จะกำหนดค่าให้น้อย ๆ โดยจากตัวอย่าง เราจะกำหนดให้มีการ mutation 2% ของข้อมูลทั้งหมด โดยวัดจากข้อมูลเป็นจำนวน bit ดังนี้

จำนวน bit = 8*5 = 40

จำนวน bit ที่จะทำ mutation = 40*0.02 = 0.8 หรือ 1 bit นั่นเอง

จากนั้น เราจะสุ่มเลือกมา 1 โครโมโซม เพื่อที่จะทำการ mutation 1 bit สมมติให้เป็นโครโมโซม V'5 ตำแหน่งที่สอง

V'5	(11010)

จะได้ค่า V''5 ที่ผ่านการ mutation ดังนี้

v''5	(10010)

หลังจากที่เราทำการ crossover และ mutation แล้ว เราจะนำข้อมูลที่ได้ไปทำหาค่าของ probability เพื่อที่จะคัดเลือกเป็นประชากรรุ่นที่สองต่อไป โดยทำขั้นตอนซ้ำจนกว่าจะได้ค่าที่ดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การพัฒนาระบบและหน้าจอการทำงาน

5.1 สภาพแวดล้อมการพัฒนาระบบ

สภาพแวดล้อมการพัฒนาระบบประกอบด้วย 2 ส่วนหลักคือ ด้านฮาร์ดแวร์ และด้านซอฟท์แวร์

ฮาร์ดแวร์ประกอบด้วย

- เครื่องคอมพิวเตอร์ Pentium 1.73 GHz
- RAM 1.5 GB
- ฮาร์ดดิสก์
- Mouse, Keyboard
- CD ROM

ซอฟท์แวร์ประกอบด้วย

- Windows XP Professional version 2002
- Microsoft visual C# 2005 express edition

5.2 การจัดเตรียมสภาพแวดล้อมในการทำงาน

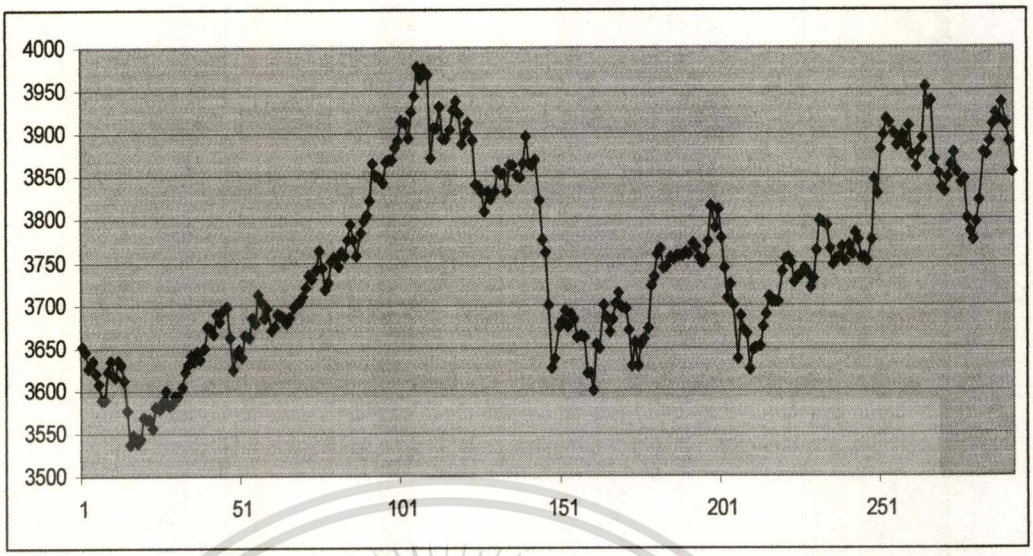
ในการพัฒนาระบบการจัดเตรียมสภาพแวดล้อมให้พร้อมสำหรับการพัฒนา โดยการติดตั้ง Window XP หลังจากนั้นจึงติดตั้ง Microsoft visual C# 2005 express edition สำหรับใช้ในการพัฒนาโปรแกรม

5.3 ขั้นตอนการทำงาน

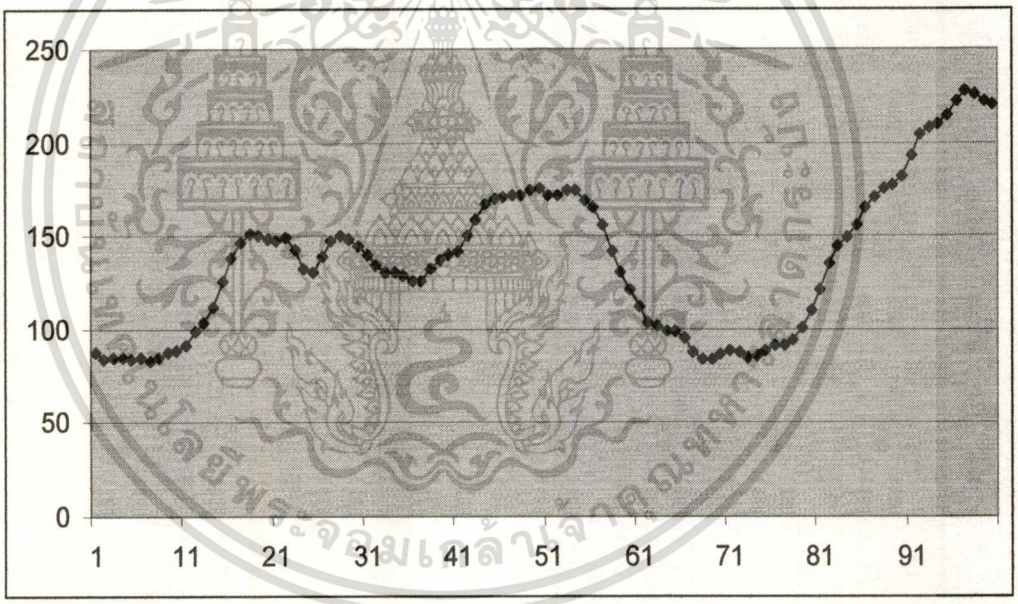
5.3.1 การเตรียมข้อมูล

การทำงานของระบบนี้ จะใช้พยากรณ์ข้อมูลใด ๆ ที่มีลักษณะเป็น Non-seasonal โดยในรายงานนี้ จะใช้ตัวอย่างของข้อมูล Dow-Jones index และ Number of internet users ซึ่งจากข้อมูลจะสร้างกราฟได้ดังรูปที่ 5.1 และ 5.2 ซึ่งจะเห็นได้ว่า ข้อมูลยังไม่มีความเป็น stationary คือมีการกวัดแกว่งของข้อมูลที่มีแนวโน้มเพิ่มขึ้น และลดลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.1 แสดงข้อมูล Dow-Jones index



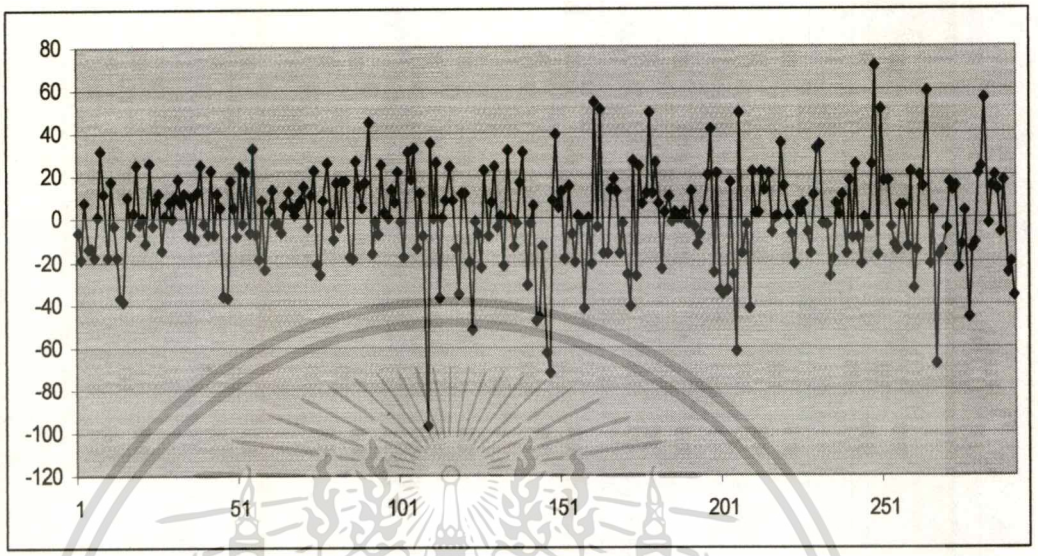
รูปที่ 5.2 แสดงข้อมูล Number of internet users

เราสามารถแปลงข้อมูลให้มีความเป็น stationary โดยใช้การ differencing ดังสมการ

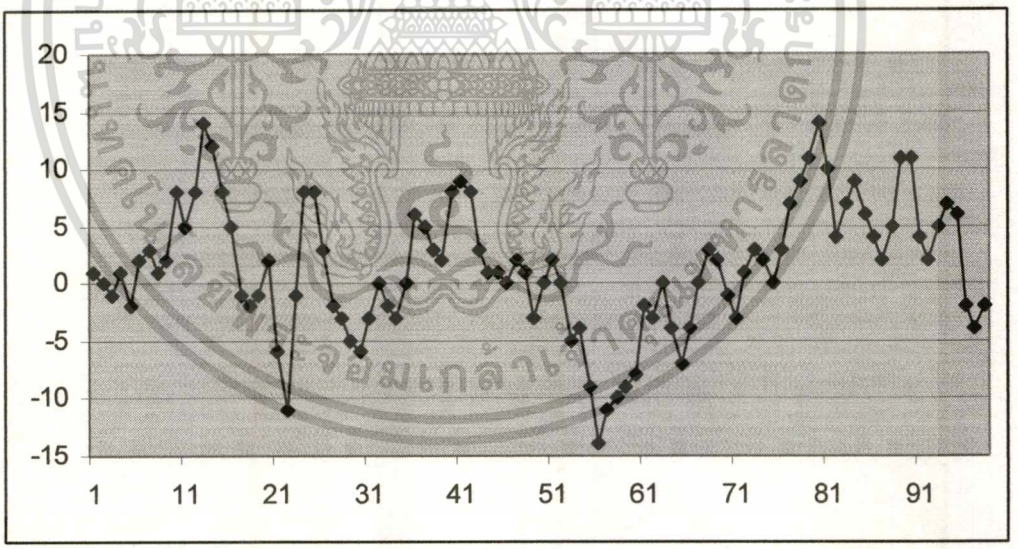
$$Y' = Y_t - Y_{t-1} \tag{5.1}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งวิธีการนี้เราจะเรียกว่า first-order differences จากข้อมูล Dow-Jones index และ Number of internet users เมื่อผ่านการทำ first-order differences จะได้ข้อมูลที่เป็น stationary ดังรูปที่ 5.2



รูปที่ 5.3 Dow-Jones index ที่ผ่านการทำ first-order differencing แล้ว



รูปที่ 5.4 Number of internet users ที่ผ่านการทำ first-order differencing แล้ว

เมื่อเราได้ differencing order ที่ทำให้ข้อมูลเป็น stationary แล้ว เราจะนำ order ที่ได้มาใช้เป็น input ของโปรแกรมต่อไป

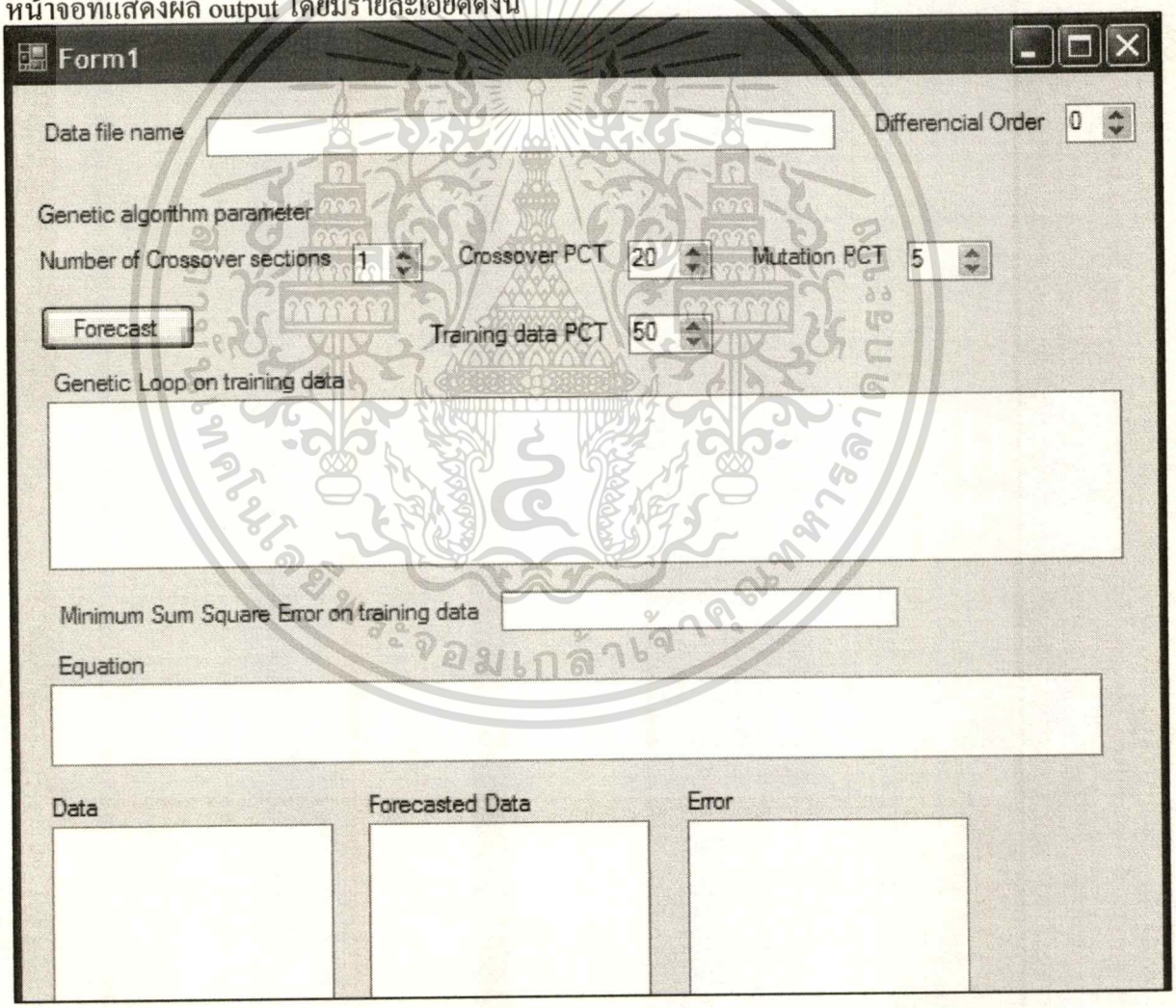
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.2 การประมวลผล

เมื่อได้ข้อมูลและ differencing order มาแล้วในส่วนของโปรแกรมจะนำเอาสมการของ autoregressive และส่วนหนึ่งของข้อมูลมาหาค่าของ parameter เพื่อนำไปสร้างสมการในการพยากรณ์ เมื่อได้สมการแล้ว เราจึงนำไปประมวลผลด้วย genetic algorithm โดยค่าที่ใช้คัดเลือกโครโมโซมในแต่ละรุ่นคือค่าของ sum square error (SSE) โดยเลือกจากโครโมโซมที่มีค่า SSE ที่มีค่าน้อยที่สุดนั่นเอง

5.4 การพัฒนาระบบและหน้าจอการทำงาน

รูปที่ 5.5 แสดงให้เห็นถึงหน้าจอของโปรแกรม ซึ่งประกอบไปด้วยหน้าจอ input และหน้าจอที่แสดงผล output โดยมีรายละเอียดดังนี้



รูปที่ 5.5 หน้าจอการประมวลผลและแสดงผลของโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4.1 หน้าจอส่วนที่รับ input และประมวลผล

หน้าจอส่วนที่รับ input เพื่อนำมาประมวลผล ประกอบด้วยส่วนต่าง ๆ ดังนี้

Data file name เป็นส่วนที่ใส่ชื่อของ text file ที่มีข้อมูลที่ต้องการพยากรณ์อยู่

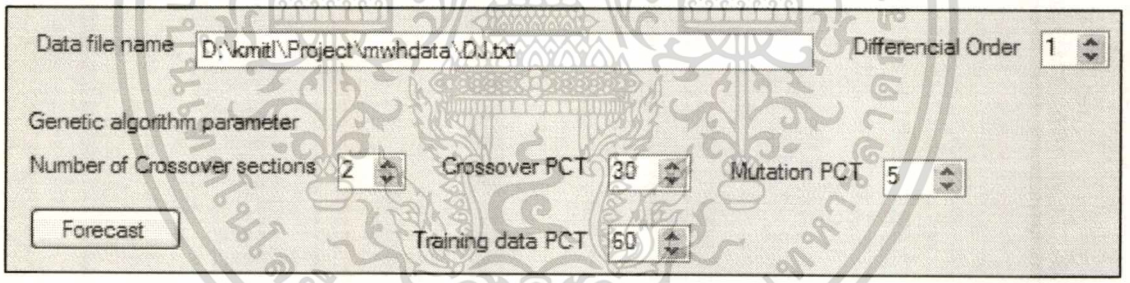
Difference order เป็นส่วนที่กำหนด order ของการ differencing ที่ทำให้ข้อมูลมีความเป็น stationary

Number of crossover sections เป็นส่วนที่กำหนดจำนวนตำแหน่งของการ crossover ในแต่ละโครโมโซม

Crossover PCT เป็นส่วนที่กำหนดเปอร์เซ็นต์ของจำนวนโครโมโซมที่นำไป crossover ในแต่ละรอบของ genetic algorithm

Mutation PCT เป็นส่วนที่กำหนดของเปอร์เซ็นต์ของจำนวนบิททั้งหมดจากทุกโครโมโซมที่นำมาเข้ากระบวนการ mutation ในแต่ละรอบของ genetic algorithm

Training data PCT เป็นส่วนที่กำหนดเปอร์เซ็นต์ของจำนวนข้อมูลที่นำมาใช้ในการสร้างสมการ autoregressive เพื่อที่จะนำมาใช้พยากรณ์



รูปที่ 5.6 หน้าจอส่วนที่รับ input และประมวลผล

จากรูปที่ 5.6 จะเห็นว่า โปรแกรมรับข้อมูล Dow-Jones index จาก file DJ.txt และเรากำหนดให้ differencing order เป็น 1 จากการหาความเป็น stationary ของข้อมูล ในส่วนที่เป็น Genetic algorithm parameter ผู้ใช้ก็สามารถกำหนดค่า parameter ได้เองเช่นเดียวกับค่าของ differencing order

ในส่วนของ Training data PCT เมื่อผู้ใช้กำหนดค่าตัวเลขแล้ว เปอร์เซ็นต์ส่วนที่เหลือจะเป็นค่าของข้อมูลที่ถูกพยากรณ์ออกมา ดังตัวอย่าง มีการกำหนดค่า training data เป็น 60 เปอร์เซ็นต์ โปรแกรมจะนำข้อมูลส่วนแรก 60 เปอร์เซ็นต์มาใช้ในการสร้างสมการ autoregressive และทำนายข้อมูลในส่วนของ 40 เปอร์เซ็นต์ที่เหลือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4.1 หน้าจอส่วนที่แสดง output

Genetic loop on training data : เป็นส่วนที่แสดงให้เห็นถึงสมการที่ให้ค่าการทำนายดีที่สุด คือมีค่า sum square error (SSE) น้อยที่สุดในแต่ละรอบของการคำนวณ genetic algorithm

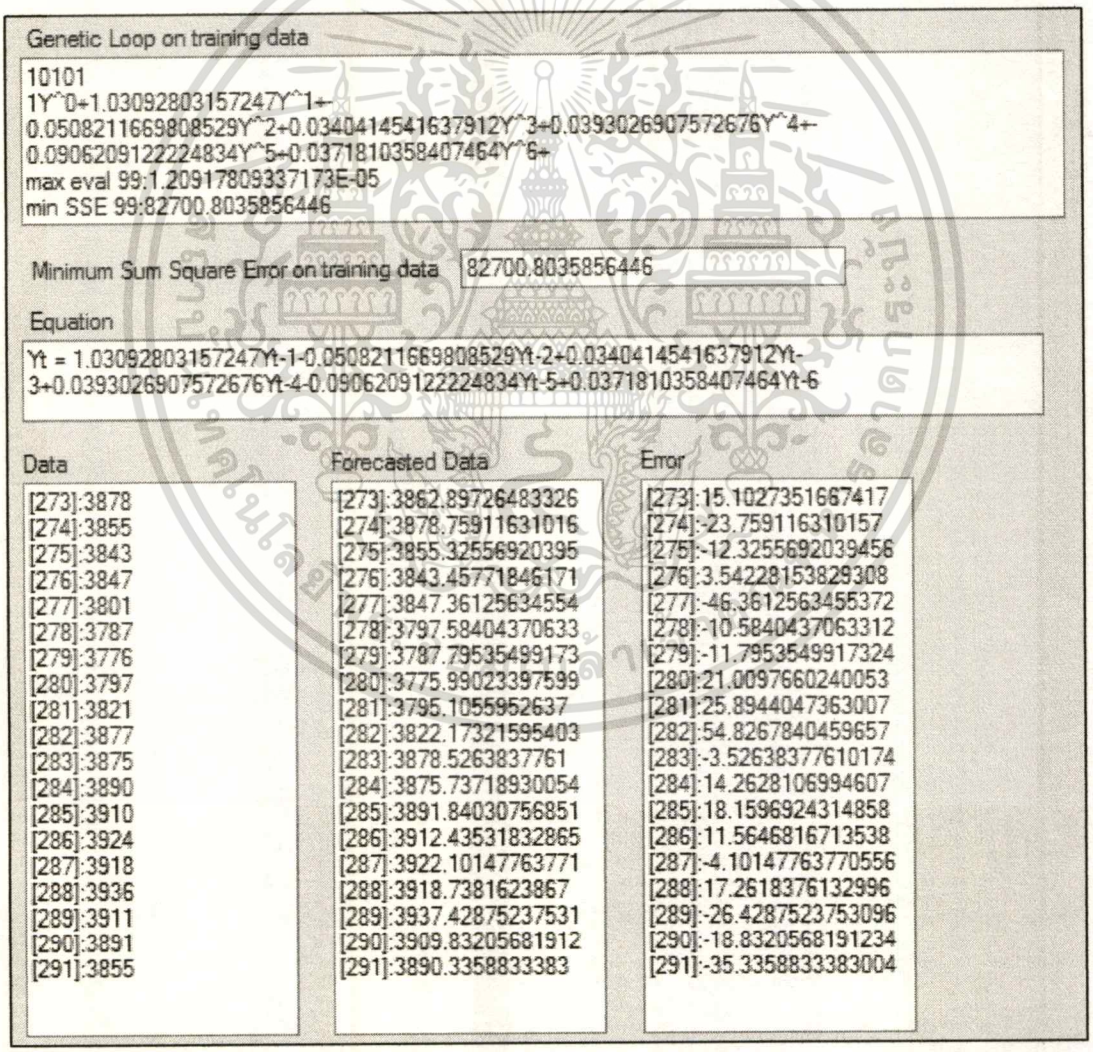
Minimum Sum Square Error on training data : เป็นส่วนที่แสดงค่าของ SSE ที่ดีที่สุดจากชุดข้อมูลที่นำมาเป็น training data

Equation : เป็นส่วนที่แสดงสมการ autoregressive ที่ได้เมื่อเสร็จสิ้นการประมวลผล

Data : เป็นส่วนที่แสดงข้อมูลจริงทั้งหมด

Forecasted Data : เป็นส่วนที่แสดงข้อมูลที่ได้จากการพยากรณ์

Error : เป็นส่วนที่แสดงค่าที่คลาดเคลื่อนจากการพยากรณ์เปรียบเทียบกับค่าจริง



รูปที่ 5.7 หน้าจอแสดง output

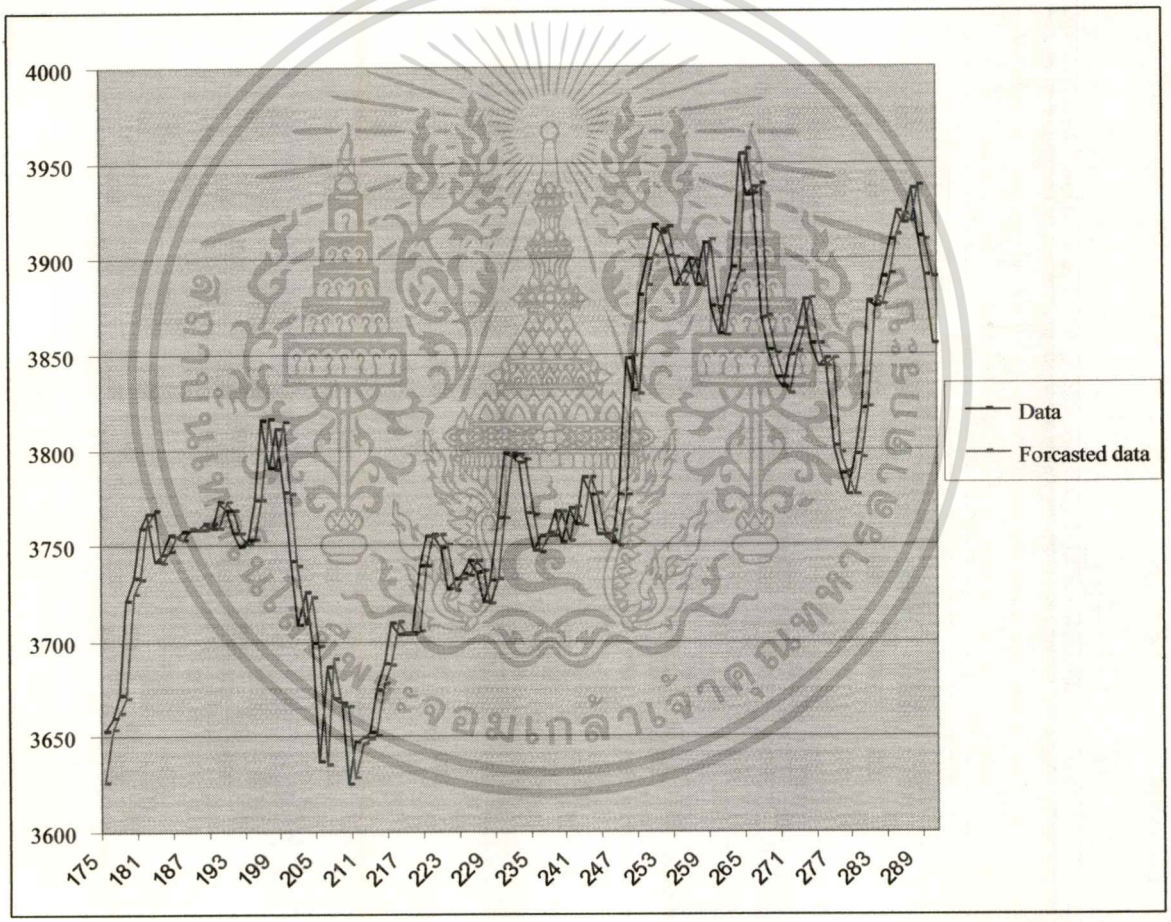
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4 ผลการทดสอบโปรแกรม

จากการทดสอบโปรแกรมกับข้อมูล Dow-Jones index จำนวน 290 ค่า จะได้สมการของ autoregressive การทำนายที่ดีที่สุดสำหรับข้อมูลนี้ ดังนี้

$$Y_t = 1.0309Y_{t-1} - 0.0508Y_{t-2} + 0.0340Y_{t-3} + 0.0393Y_{t-4} - 0.0906Y_{t-5} + 0.0376Y_{t-6} + e_t$$

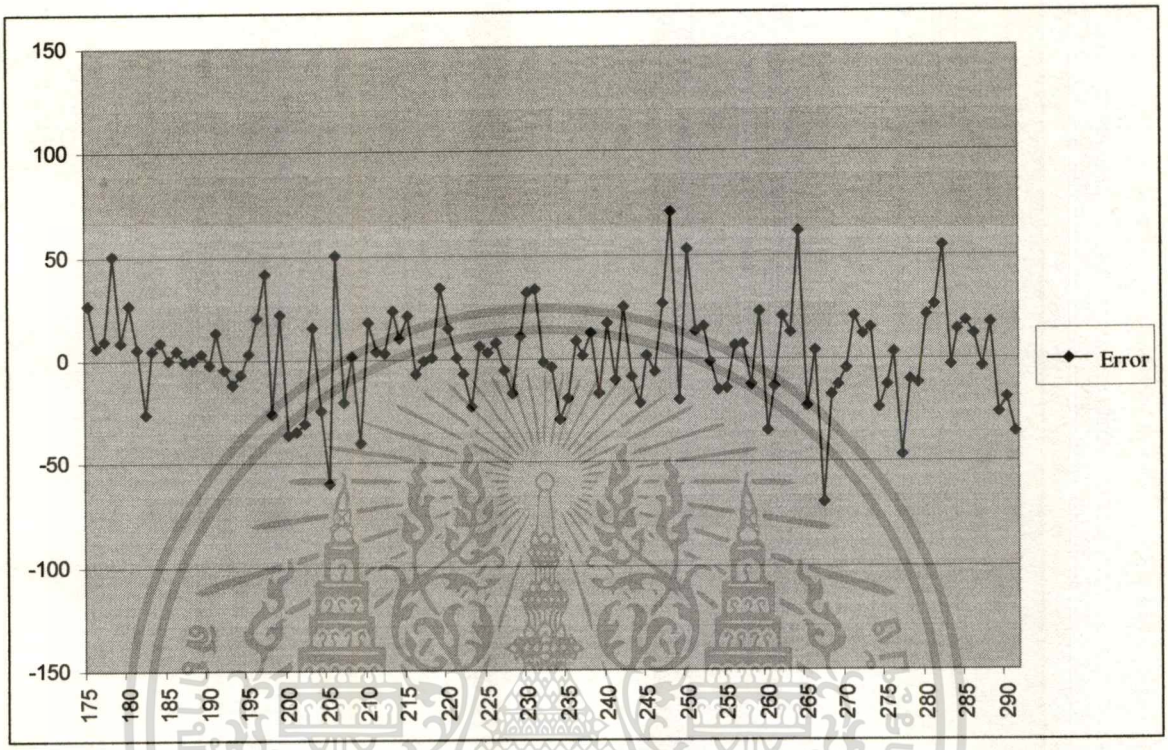
ค่าของข้อมูลจริงและค่าของข้อมูลที่ทำนายได้ จะนำมาสร้างเป็นกราฟได้ดังรูปที่ 5.6 โดยเมื่อเราใส่ค่า Training data PCT เป็น 60 เปอร์เซ็นต์ โปรแกรมจะพยากรณ์ข้อมูล 40 เปอร์เซ็นต์ที่เหลือ



รูปที่ 5.8 แสดงผลของค่า Dow-Jones index ที่พยากรณ์เปรียบเทียบกับค่าจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากข้อมูลของการพยากรณ์ค่า Dow-Jones index จะได้ค่าของความความคลาดเคลื่อนดังกราฟในรูป 5.9



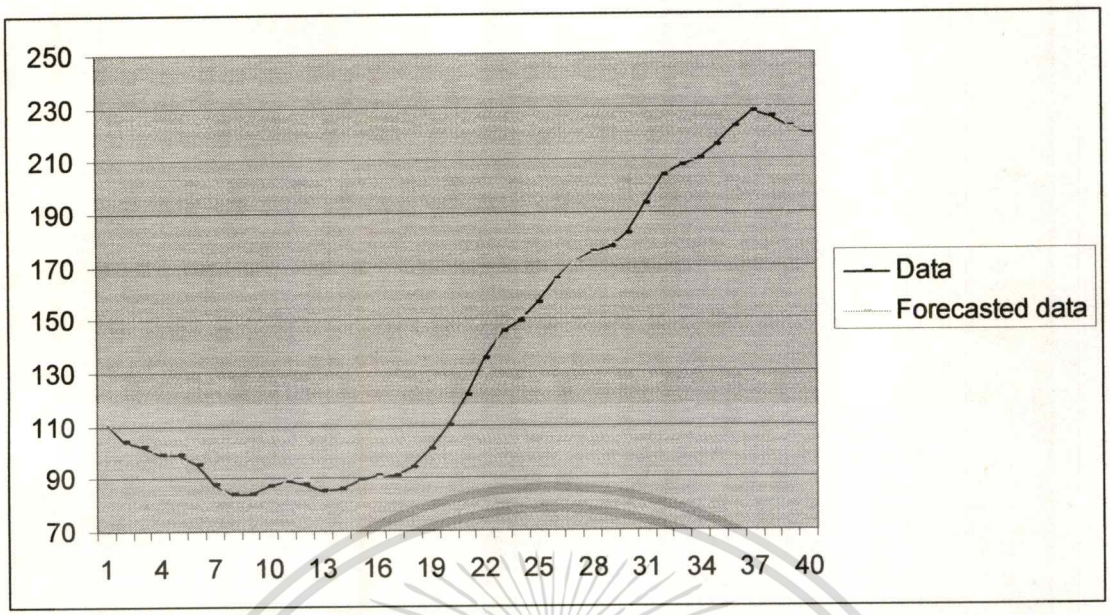
รูปที่ 5.9 ค่าความคาดเคลื่อนที่ได้จากการพยากรณ์ค่า Dow-Jones index

จากการทดสอบโปรแกรมกับข้อมูล Number of internet users จำนวน 100 ค่า จะได้สมการของautoregressive การทำนายที่ดีที่สุดสำหรับข้อมูลนี้ ดังนี้

$$Y_t = 2.1803Y_{t-1} - 1.88078Y_{t-2} + 1.1529Y_{t-3} + 0.5948Y_{t-4} - 0.1409Y_{t-5} + e_t$$

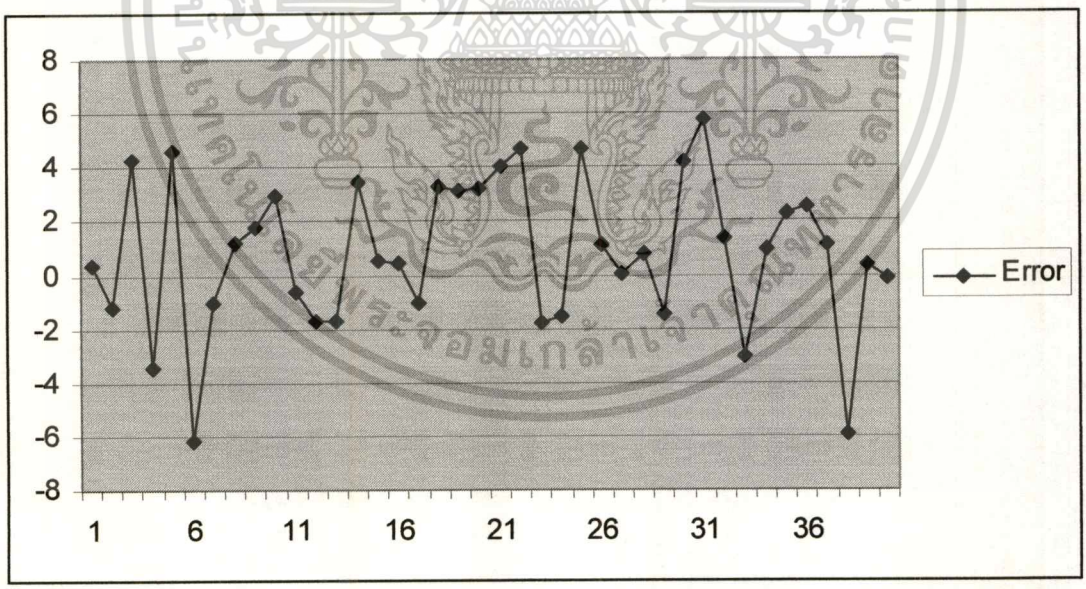
ค่าของข้อมูลจริงและค่าของข้อมูลที่ทำนายได้ จะนำมาสร้างเป็นกราฟได้ดังรูปที่ 5.10 โดยเมื่อเราใส่ค่า Training data PCT เป็น 40 เปอร์เซ็นต์ โปรแกรมจะพยากรณ์ข้อมูล 60 เปอร์เซ็นต์ที่เหลือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.10 แสดงผลของค่า Number of internet users ที่พยากรณ์เปรียบเทียบกับค่าจริง

จากข้อมูลของการพยากรณ์ค่า Number of internet users จะได้ค่าของความความคลาดเคลื่อนดังกราฟในรูป 5.11



รูปที่ 5.7 ค่าความคาดเคลื่อนที่ได้จากการพยากรณ์ค่า Number of internet users

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผลการพัฒนาระบบและข้อเสนอแนะ

6.1 ผลการวิเคราะห์และออกแบบระบบ

ระบบช่วยการพยากรณ์แบบ Autoregressive โดยใช้วิธี Genetic algorithm ทำให้ช่วยค้นหา model ของการพยากรณ์ได้รวดเร็วและแม่นยำขึ้น โดยในส่วนของ Genetic algorithm จะเป็นส่วนที่ช่วยในการคัดเลือก model ที่ดีที่สุดของการทำ autoregressive ขึ้นมาแล้วใช้ model ที่ถูกคัดเลือกมาทำการพยากรณ์ โดยในส่วนของโปรแกรมจะให้ผู้ใช้สามารถปรับค่า parameter ต่าง ๆ ในส่วนของของdifferencing order , genetic algorithm และจำนวนของข้อมูลที่จะนำไปสร้าง autoregressive model ได้

ในส่วนของการแสดงผลของข้อมูล มีการแสดงผลของสมการที่ใช้ในการพยากรณ์ ,ข้อมูลที่ถูกพยากรณ์ และค่าความคลาดเคลื่อนของการพยากรณ์ ซึ่งจะทำผู้ใช้ให้นำข้อมูลไปวิเคราะห์ได้สะดวก

6.2 ประโยชน์ที่คาดว่าจะได้รับ

สำหรับการพัฒนาระบบนี้ จะทำให้ผู้ใช้งานสามารถนำระบบไปใช้พยากรณ์ข้อมูลในลักษณะที่เป็น non-seasonal ได้อย่างมีประสิทธิภาพโดยไม่ต้องกำหนด model ในการทำ auto regressive เอง ทำให้การพยากรณ์ทำได้อย่างสะดวก และรวดเร็ว เนื่องจากโปรแกรมมีการคัดเลือก model ในการใช้งานไว้ให้แล้ว

6.3 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต

- พัฒนาเพิ่มเติมในส่วนของ moving average model เพื่อนำไปใช้พยากรณ์ในรูปแบบของ ARIMA ได้
- พัฒนาหน้าจอการแสดงผลเพิ่มในส่วนของกราฟ เพื่อให้ง่ายต่อการวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

ชนศักดิ์ บ่ายเที่ยง และ ศรีบุตร แววเจริญ. 2542. **คณิตศาสตร์วิศวกรรมและวิทยาศาสตร์.** ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Chorng-Shyong Ong, Jih-Jeng Huang,Gwo-Hshiung Tzeng. 2005. **Model identification of ARIMA family using genetic algorithms.** [Online], Available: www.elsevier.com/locate/amc

Engineering fundamental. 2008. **Least-Square Multiple Regression.** [Online]. Available: http://www.efunda.com/math/leastsquares/lstsqrzrwtxy1d.cfm

Middle Tennessee state university. **Regression tutorial project.** [Online]. Available: http://mtsu32.mtsu.edu:11308/regression.htm

Spyros Makridakis, Steven C. Wheelwrigth, Rob J. Hyndman. 1997. **Forecasting methods and applications.** Newyork: John Wiley & Sons Inc.

Zbigniew Michalewicz. 1996. **Genetic algorithms + data structures = evolution programs.** Berlin : Springer-Verlag

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นางสาวสายสุดา ธนาธรรมนันท์
วัน เดือน ปีเกิด	17 ธันวาคม 2520
สถานที่เกิด	จังหวัดกรุงเทพมหานคร
ประวัติการศึกษา	ระดับชั้นประถมศึกษา โรงเรียนแสงหิรัญ จ.กรุงเทพมหานคร
	โรงเรียนศรีวิกรม์ จ.กรุงเทพมหานคร
	ระดับชั้นมัธยมศึกษา โรงเรียนศรีวิกรม์ จ.กรุงเทพมหานคร
	ระดับอุดมศึกษา วิศวกรรมศาสตร์บัณฑิต สาขาวิชคอมพิวเตอร์
	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ประสบการณ์การทำงาน	บริษัท คอมพิวเตอร์ยูเนียน จำกัด (พ.ศ. 254 3 - พ.ศ.2545)
	บริษัท พาราแอดวานซ์อินโฟเทค จำกัด (พ.ศ. 2545 - ปัจจุบัน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้