การเปรียบเทียบวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียม และวิธีจีเนติกอัลกอริทึม โดยใช้องค์ประกอบหลักในการพยากรณ์อินทรีย์วัตถุในดิน : กรณีศึกษาพื้นที่เกษตรกรรมสวนผลไม้ในภาคตะวันตกของประเทศไทย

โดย

นางสาวสิริกัลยา ประมวล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์

ภาควิชาสถิติ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2552

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

การเปรียบเทียบวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียม และวิธีจีเนติกอัลกอริทึม โดยใช้องค์ประกอบหลักในการพยากรณ์อินทรีย์วัตถุในดิน : กรณีศึกษาพื้นที่เกษตรกรรมสวนผลไม้ในภาคตะวันตกของประเทศไทย

โดย

นางสาวสิริกัลยา ประมวล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์

ภาควิชาสถิติ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2552

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

COMPARISONS OF MULTIPLE LINEAR REGRESSION, ARTIFICIAL NEURAL NETWORKS AND GENETIC ALGORITHM BASED ON PRINCIPAL COMPONENTS FOR PREDICTING SOIL ORGANIC MATTER: A CASE STUDY FROM FRUIT FARMING AGRICULTURAL LOCATIONS IN WESTERN REGION OF THAILAND

By

Sirikanlaya Pramual

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

Department of Statistics

Graduate School

SILPAKORN UNIVERSITY

2009

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร อนุมัติให้วิทยานิพนธ์เรื่อง “การเปรียบเทียบวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมและวิธีจีเนติกอัลกอริทึม โดยใช้องค์ประกอบหลักในการพยากรณ์อินทรีย์วัตถุในดิน: กรณีศึกษาพื้นที่เกษตรกรรมสวนผลไม้ในภาคตะวันตกของประเทศไทย” เสนอโดย นางสาว สิริกัลยา ประมวล เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์

……..
(รองศาสตราจารย์ ดร.ศิริชัย ชินะตังกูร)
คณบดีบัณฑิตวิทยาลัย
วันที่.....เดือน..... พ.ศ.....

อาจารย์ที่ปรึกษาวิทยานิพนธ์
ผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ

คณะกรรมการตรวจสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.กุศยา ปลั่งพงษ์พันธ์)
...../...../.....

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โฉมที)
...../...../.....

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ)
...../...../.....

50304207 : สาขาวิชาสถิติประยุกต์
คำสำคัญ : อินทรีย์วัตถุในดิน/การวิเคราะห์องค์ประกอบหลัก/การถดถอยเชิงเส้นพหุคูณ/โครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า/จีเนติกอัลกอริทึม

สิริกัลยา ประมวล : การเปรียบเทียบวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมและวิธีจีเนติกอัลกอริทึม โดยใช้องค์ประกอบหลักในการพยากรณ์อินทรีย์วัตถุในดิน : กรณีศึกษาพื้นที่เกษตรกรรมสวนผลไม้ในภาคตะวันตกของประเทศไทย. อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ผศ.ดร.กมลชนก พานิชการ. 81 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบ วิธีการสำหรับการพยากรณ์ปริมาณอินทรีย์วัตถุในดินสามวิธี คือวิธีการถดถอยเชิงเส้นพหุคูณ วิธีการโครงข่ายประสาทเทียมแบบเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม โดยเก็บตัวอย่างดินมาจากพื้นที่สวนผลไม้ในภาคตะวันตกของประเทศไทย 3 จังหวัดได้แก่ นครปฐม สมุทรสาคร และสมุทรสงคราม วัดข้อมูลเกี่ยวกับคุณสมบัติทางเคมีของดิน แล้วใช้เป็นตัวแปรพยากรณ์ทั้งหมด 17 ตัวแปร วิธีวิเคราะห์องค์ประกอบหลักถูกนำมาใช้ ทำให้ได้จำนวนตัวแปรที่เหมาะสมในโมเดล เป็นการลดมิติข้อมูลและขจัดความสัมพันธ์ระหว่างตัวแปรพยากรณ์ ซึ่งผลลัพธ์สามารถลดจำนวนตัวแปรเหลือเพียง 5 องค์ประกอบหลัก ได้แก่ PC1, PC2, PC3, PC4 และ PC5 สามารถอธิบายความผันแปรของข้อมูลได้ 75.81% องค์ประกอบหลักทั้ง 5 ถูกใช้เป็นตัวแปรในการพยากรณ์ปริมาณอินทรีย์วัตถุ ในการเปรียบเทียบประสิทธิภาพของโมเดลใช้ค่า AI, RMSE, MBE และ MAE

ผลที่ได้จากงานวิจัยพบว่าผลลัพธ์ที่ได้จากวิธีการถดถอยเชิงเส้นพหุคูณมีค่าวัดประสิทธิภาพเท่ากับ 0.97, 0.27, -0.11 และ 0.23 ตามลำดับ ผลลัพธ์ที่ได้จากวิธีการโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า มีค่าวัดประสิทธิภาพเท่ากับ 0.99, 0.24, -0.17 และ 0.24 ตามลำดับ และผลลัพธ์ที่ได้จากวิธีจีเนติกอัลกอริทึม มีค่าวัดประสิทธิภาพเท่ากับ 0.99, 0.11, -0.03 และ 0.09 ตามลำดับ แสดงให้เห็นว่าประสิทธิภาพของการพยากรณ์ได้จากวิธีการจีเนติกอัลกอริทึมดีกว่า การพยากรณ์ที่ได้จากวิธีการถดถอยเชิงเส้นพหุคูณและวิธีการโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า

50304207 : MAJOR : APPLIED STATISTICS
KEY WORDS : SOIL ORGANIC MATTER/PRINCIPAL COMPONENT ANALYSIS/MULTIPLE LINEAR REGRESSION/FEED-FORWARD ARTIFICIAL NEURAL NETWORK/GENETIC ALGORITHM
SIRIKANLAYA PRAMUAL : COMPARISONS OF MULTIPLE LINEAR REGRESSION ARTIFICIAL NEURAL NETWORKS AND GENETIC ALGORITHM BASED ON PRINCIPAL COMPONENTS FOR PREDICTING SOIL ORGANIC MATTER : A CASE STUDY FROM FRUIT FARMING AGRICULTURAL LOCATIONS IN WESTERN REGION OF THAILAND. THESIS ADVISOR : ASST.PROF.KAMOLCHANOK PANISHKAN, Ph.D. 81 pp.

The objective of this research is to study 3 methods; namely multiple linear regression, feed-forward artificial neural network and genetic algorithm for predicting quantity of organic matter. Soil samples were selected from fruit farming agricultural area in Nakhon Pathom, Samut Sakhon and Samut Songkram. Soil samples were measured 17 soil properties used as independent variables. In order to reduce numbers of independent variables and eliminate data multicolinearity, Principal Component Analysis were used. The models with 5 principal components such as PC1, PC2, PC3, PC4 and PC5 accounted for 75.8% of total variance were performed. The evaluations of performance of models were measured by performance indexes which are AI, RMSE, MBE and MAE.

The performance indexes from multiple linear regression model are 0.97, 0.27, -0.11 and 0.23, respectively. The performance indexes from feed-forward artificial neural network model are 0.99, 0.24, -0.17 and 0.22, respectively. The performance indexes from genetic algorithm model are 0.99, 0.11, -0.03 and 0.09, respectively. The result show that genetic algorithm model is better than multiple linear regression and feed-forward artificial neural network model for predicting soil organic matter.

Department of Statistics Graduate School, Silpakorn University Academic Year 2009
Student's signature .....
Thesis Advisor's signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี โดยได้รับความอนุเคราะห์จากอาจารย์หลายๆ ท่าน ผู้วิจัยขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษา แนะนำ และแก้ไขข้อบกพร่องต่างๆ ในงานวิจัยของผู้วิจัยมาโดยตลอด พร้อมทั้งขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.กุศยา ปลั่งพงษ์พันธ์ ที่กรุณาเป็นประธานกรรมการในการสอบ และผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โฉมที ที่ได้กรุณาเป็นผู้ทรงคุณวุฒิ ที่ให้คำแนะนำตรวจสอบ คำชี้แนะและมุมมองที่ดีที่ทำให้การดำเนินงานวิจัยประสบผลสำเร็จและทำให้วิทยานิพนธ์มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.นัทธีรา สรรมณี อาจารย์ภาควิชาวิทยาศาสตร์สิ่งแวดล้อม คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ที่ให้ความอนุเคราะห์ข้อมูลจากโครงการ “บทบาทของกรดฮิวมิกและกรดฟัลวิกในดินของพื้นที่เกษตรกรรมต่อการดูดซับจุลธาตุอาหารในดินและโลหะที่เป็นพิษ: กรณีศึกษาภูมิภาคตะวันตกของประเทศไทย” ในการทำงานวิจัยครั้งนี้

ผู้วิจัยขอขอบพระคุณคณาจารย์ ภาควิชาสถิติทุกท่านที่ได้ให้ความรู้และความช่วยเหลือพร้อมทั้งให้กำลังใจ ผู้วิจัยตลอดระยะเวลาในการศึกษา ขอขอบคุณ คุณนงลักษณ์ เอี๊ยวเจริญ ที่ให้การช่วยเหลือและอำนวยความสะดวกงานเอกสารแก่ผู้วิจัยตลอดมา และขอขอบคุณ คุณพิศาล สุขขี ที่ให้คำแนะนำและความรู้ด้านการเขียนโปรแกรม รวมทั้ง พี่ๆ น้องๆ ในภาควิชาสถิติทุกคน ที่เป็นกำลังใจและช่วยเหลือ ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงด้วยดี

สุดท้ายนี้คุณค่าและคุณประโยชน์ของวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเพื่อตอบแทนพระคุณของคุณพ่อพิสิษฐ์ ประมวล คุณแม่มัลนิกา ประมวล และคุณพ่อแบน สุขขี ที่ให้การสนับสนุนและส่งเสริมการทำงานของผู้วิจัยในทุกๆ ด้าน

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญตาราง ..... ญ

สารบัญภาพ ..... ฏ

บทที่

1 บทนำ 1

ความสำคัญและความเป็นมาของปัญหา 1

วัตถุประสงค์ของการศึกษา 3

สมมติฐานของการศึกษา 4

ขอบเขตของการศึกษา 4

ขั้นตอนการศึกษา 4

ประโยชน์ที่คาดว่าจะได้รับ 5

นิยามศัพท์เฉพาะ 5

2 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง 7

การวิเคราะห์องค์ประกอบหลัก 7

หลักการของการวิเคราะห์องค์ประกอบหลัก 7

ความสัมพันธ์ระหว่างตัวแปรเดิมกับตัวประกอบหลัก 8

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ 10

วิธีโครงข่ายประสาทเทียม 11

แนวคิดพื้นฐานของโครงข่ายประสาทเทียม 11

การสร้างโมเดลโครงข่ายประสาทเทียม 13

ข้อมูลป้อนเข้า 13

ชั้นซ่อน 13

ค่าถ่วงน้ำหนักและไบแอส 14

ฟังก์ชั่นการรวมและฟังก์ชั่นการแปลงค่า 15

กระบวนการสอนหรือการเรียนรู้ 15

การหยุดสอน 20

บทที่ หน้า
ชุดข้อมูล 21
การทดสอบโครงข่าย 21
วิธีจีเนติกอัลกอริทึม 21
ขั้นตอนการทำงานของจีเนติกอัลกอริทึม 22
องค์ประกอบที่สำคัญของจีเนติกอัลกอริทึม 23
การเข้ารหัสโครโมโซม 23
ประชากรต้นแบบ 25
ฟังก์ชั่นวัตถุประสงค์ 25
ปฏิบัติการทางสายพันธุ์ 25
พารามิเตอร์ 29
เงื่อนไขในการหยุดกระบวนการค้นหาคำตอบ 29
งานวิจัยที่เกี่ยวข้อง 30
3 วิธีดำเนินงานวิจัย 36
ขั้นตอนการดำเนินงานวิจัย 36
ตัวแปรที่ใช้ในการวิจัย 37
การสร้างโมเดลด้วยวิธีการถดถอยเชิงเส้นพหุคูณ 39
การสร้างโมเดลด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า . 40
การสร้างโมเดลด้วยวิธีจีเนติกอัลกอริทึม 43
เครื่องมือที่ใช้ในการวิเคราะห์ 47
4 ผลการวิจัย 48
ส่วนที่ 1 การวิเคราะห์ข้อมูลคุณสมบัติทางเคมีของดินและการวิเคราะห์ .....
องค์ประกอบหลัก 48
ส่วนที่ 2 ผลลัพธ์ที่ได้จากการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน 52
ผลลัพธ์จากวิธีการถดถอยเชิงเส้นพหุคูณ 52
ผลลัพธ์จากวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า 54
ผลลัพธ์จากวิธีจีเนติกอัลกอริทึม 57
การเปรียบเทียบประสิทธิภาพของการพยากรณ์จากทั้งสามวิธี 59
5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 62
สรุปผลการวิเคราะห์องค์ประกอบหลัก 62

บทที่ หน้า
สรุปผลการเปรียบเทียบประสิทธิภาพจากการพยากรณ์ 63
อภิปรายผล 64
ข้อจำกัดในการวิจัย..... 65
ข้อเสนอแนะจากงานวิจัย..... 66

บรรณานุกรม..... 66

ภาคผนวก 70
ภาคผนวก ก โปรแกรมการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม 71

ประวัติผู้วิจัย 81

สารบัญตาราง

ตารางที่		หน้า
1	ค่าต่ำสุดและค่าสูงสุดของสัมประสิทธิ์การถดถอยที่ได้จากการสร้างช่วงความเชื่อมั่น 95%	45
2	ค่าสถิติพื้นฐานของคุณสมบัติทางเคมีของตัวอย่างดินที่นำมาศึกษา	49
3	ค่าน้ำหนักของ 5 องค์ประกอบหลักแรกของคุณสมบัติทางเคมีของดิน	51
4	ค่าประมาณของพารามิเตอร์จากวิธีการถดถอยเชิงเส้นพหุคูณ	52
5	ปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ	53
6	ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ	54
7	ปริมาณอินทรีย์วัตถุจากการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า	55
8	ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีการโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า	57
9	ค่าสัมประสิทธิ์ของสมการการถดถอยเชิงเส้นพหุคูณ	57
10	ปริมาณอินทรีย์วัตถุจากการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม	58
11	ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม	59
12	ปริมาณอินทรียวัตถุจากการพยากรณ์จากทั้ง 3 โมเดล	60
13	ค่าวัดประสิทธิภาพการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ (MLR) วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า (FANN) และวิธีจีเนติกอัลกอริทึม(GA)	61

สารบัญภาพ

ภาพที่		หน้า
1	โครงสร้างระบบประสาทในสมอง	12
2	กระบวนการทำงานของโครงข่ายประสาทเทียม	15
3	การเรียนรู้แบบไม่มีการเรียนการสอน (Unsupervised Learning)	16
4	การเรียนรู้แบบมีการเรียนการสอน (Supervised Learning)	17
5	โครงข่ายประสาทเทียมแบบแพร่กลับ	19
6	กรณี Over fitting	20
7	การเข้ารหัสโครโมโซมแบบต้นไม้	25
8	ตัวอย่างของ One-point crossover	27
9	ตัวอย่างของ Two-point crossover	28
10	ตัวอย่างของการมิวเทชั่น	28
11	ภาพรวมขั้นตอนการศึกษา	37
12	ขั้นตอนการสร้างโมเดลการถดถอยเชิงเส้นพหุคูณ	40
13	ขั้นตอนการสร้างโมเดล FANN	42
14	รูปแบบการสร้างโครโมโซม	43
15	กระบวนการทำงานของจีเนติกอัลกอริทึมที่มีการแทนที่แบบคงตัว	46
16	ขั้นตอนการทำงานของจีเนติกอัลกอริทึม	47
17	Scree plot สำหรับการเลือกองค์ประกอบหลัก	50
18	การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี MLR ของชุดข้อมูลทดสอบ	53
19	ค่าน้ำหนักของโมเดลที่ดีที่สุดของการเรียนรู้โครงข่ายด้วยวิธี FANN	55
20	การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี FANN ของชุดข้อมูลทดสอบ	56
21	การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี GA ของชุดข้อมูลทดสอบ	58
22	การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี MLR, ANN และ GA ของชุดข้อมูลทดสอบ	60
23	ระดับค่าวัดประสิทธิภาพของการพยากรณ์ด้วยวิธี MLR, FANN และ GA	63

บทที่ 1
บทนำ

ความสำคัญและความเป็นมาของปัญหา

การสร้างโมเดลทางคณิตศาสตร์เพื่อการพยากรณ์ที่มีประสิทธิภาพ เป็นที่นิยมมากในปัจจุบัน เนื่องจากทำให้เราสามารถที่จะคาดการณ์สถานการณ์ล่วงหน้าในเรื่องนั้นๆ เพื่อที่จะหาทางป้องกันผลเสียที่อาจจะเกิดขึ้นในอนาคต การสร้างโมเดลสำหรับการพยากรณ์เป็นที่นิยมในศาสตร์หลายด้าน ไม่ว่าจะเป็นในด้านสิ่งแวดล้อม เช่น การพยากรณ์ความเข้มข้นของโอโซน (Sousa et al. 2007 : 97-103) เพื่อใช้เตือนภัยจากระดับความเข้มข้นของโอโซน การพยากรณ์จำนวนการเกิดไฟไหม้ (Yang et al. 2006 : 213-219) เพื่อการวางแผนการจัดการและลดความสูญเสียจากการเกิดไฟไหม้ ในด้านการเกษตรใช้ในการพยากรณ์คาร์บอนอินทรีย์ (Ingleby and Crowe 2000 : 57-63) หรือการพยากรณ์อินทรีย์วัตถุ (Ingleby and Crowe 2001 : 7.1-7.5) ที่เป็นตัวชี้วัดคุณภาพของดิน เพื่อประโยชน์ในการวางแผนและปรับปรุงคุณภาพของดิน ในด้านการแพทย์ (Jung Yi Kim et al. 2005 : 6942-6945, Song et al. 2005 : 487-495) ด้านวิศวกรรมการผลิต (Erenturk 2007 : 905-912) และด้านการเงินการธนาคาร (Rurkhamet 1998 : 21-28) เป็นต้น ดังนั้นจะเห็นได้ว่าการสร้างโมเดลที่มีความแม่นยำนั้นมีความจำเป็นและมีความสำคัญมากในการพัฒนางานในด้านต่างๆ

วิธีการทางสถิติที่นิยมใช้ในการพยากรณ์ตัวแปรตามที่เป็นแบบเชิงเส้นนั้น ได้แก่วิธีการถดถอยเชิงพหุคูณ เนื่องจากเป็นวิธีที่เข้าใจง่ายและใช้งานได้ง่าย เช่นในงานวิจัยของ Ingleby และ Crowe (Ingleby and Crowe 2001 : 7.1-7.5) ที่ใช้วิธีการถดถอยเชิงพหุคูณในการพยากรณ์คาร์บอนอินทรีย์ และมีงานวิจัยด้านอื่นๆ ที่ใช้วิธีการถดถอยเชิงพหุคูณในการสร้างโมเดลในการพยากรณ์อีกหลายงานวิจัย (Eskandari, Rewaee and Mohammadnia 2004: 40-48 ; Pao 2008: 720-727) ในปัจจุบันมีวิธีการและเทคนิคใหม่ที่ได้รับความนิยมในการนำมาสร้างโมเดลในการพยากรณ์ นั่นคือวิธีการโครงข่ายประสาทเทียมและวิธีการจีเนติกอัลกอริทึม ซึ่งวิธีการโครงข่ายประสาทเทียมเป็นวิธีการหาคำตอบที่เลียนแบบมาจากการทำงานในสมองของมนุษย์ ซึ่งมีการเรียนรู้ชุดข้อมูล รวมทั้งเป็นโมเดลที่ไม่คำนึงถึงความเป็นเชิงเส้นของข้อมูล ซึ่งปัจจุบันมีการนำมาประยุกต์ในการสร้างโมเดล

สำหรับการพยากรณ์ เช่นในงานวิจัยของ Ingleby และ Crowe (Ingleby and Crowe 2001 : 57-63) ที่ใช้วิธีการโครงข่ายประสาทเทียมในการพยากรณ์อินทรีย์วัตถุ ส่วนวิธีการจีเนติกอัลกอริทึมเป็นวิธีการค้นหาคำตอบโดยใช้หลักการคัดเลือกแบบธรรมชาติและหลักการทางสายพันธุ์ เป็นการคำนวณที่มีวิวัฒนาการอยู่ในขั้นตอนของการหาคำตอบ ซึ่งปัจจุบันเป็นที่ยอมรับในประสิทธิภาพของวิธีการ (อาทิตย์ ศรีแก้ว 2002) และมีการนำไปประยุกต์กันอย่างกว้างขวาง (Karunaka and Datta 2007 : 58-66 ; Yang and Gell 2006 : 213-219) ทั้งวิธีโครงข่ายประสาทเทียมและวิธีจีเนติกอัลกอริทึม เหมาะสำหรับปัญหาที่มีการเปลี่ยนแปลงของข้อมูลอยู่ตลอดเวลา ในงานวิจัยนี้จะใช้สามวิธีดังกล่าวในการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน

เนื่องจากประเทศไทยเป็นประเทศเกษตรกรรมและส่วนใหญ่เป็นเกษตรกรรมแบบดั้งเดิม ที่มีการเพาะปลูกติดต่อกันเป็นเวลายาวนาน โดยไม่มีการปรับปรุงบำรุงดิน หรือการเติมอินทรีย์วัตถุที่ไม่มากเพียงพอ เป็นสาเหตุให้ปริมาณอินทรีย์วัตถุลดต่ำลงเรื่อยๆ (คณาจารย์ภาควิชาปฐพีวิทยา 2548) และเกิดผลเสียต่อการเกษตรในระยะยาว ดังนั้นการสร้างโมเดลในการพยากรณ์ปริมาณอินทรีย์วัตถุ จึงมีความสำคัญในด้านการเป็นเครื่องมือในการตรวจสอบคุณภาพของดิน ที่จะช่วยในการตัดสินใจและวางแผนการใช้ดินเพื่อให้เกิดประโยชน์สูงสุด และการปรับปรุงดินเพื่อพัฒนาผลผลิตทางด้านการเกษตร เนื่องจากอินทรีย์วัตถุ (Organic Matter: OM) ในดินจัดว่าเป็นแหล่งสำรองของธาตุอาหารในดินและมีส่วนสำคัญในการเสริมสร้างคุณสมบัติของดินทั้งทางกายภาพ ทางชีวภาพและทางเคมีให้เหมาะสมต่อการเจริญเติบโตของพืช บทบาทของอินทรีย์วัตถุที่มีผลต่อทางกายภาพของดิน เช่น ความหนาแน่นและความพรุนของดิน โครงสร้างของดินและความสามารถในการอุ้มน้ำของดิน บทบาททางด้านชีวภาพเช่น การแปรสภาพธาตุอาหาร การปลดปล่อยไนโตรเจนในรูปของแอมโมเนียหรือไนเตรตที่เป็นประโยชน์ รวมทั้งช่วยยับยั้งการเจริญเติบโตของเชื้อโรคในพืช (Stevenson 1994) บทบาททางด้านเคมีของดิน เช่นเป็นแหล่งธาตุอาหารให้แก่พืชเช่น ไนโตรเจน เหล็ก สังกะสี และ ทำให้ดินมีค่าการแลกเปลี่ยนประจุบวก (CEC) ในดินสูงขึ้นซึ่งจะช่วยในการดูดซับสารอาหาร ซึ่งในงานวิจัยนี้จะใช้ข้อมูลคุณสมบัติทางเคมีของดินที่มีความสัมพันธ์กับอินทรีย์วัตถุในดิน ในการพยากรณ์หาปริมาณอินทรีย์วัตถุที่เป็นคุณสมบัติอย่างหนึ่งในการบ่งชี้คุณภาพของดินได้

ดังนั้นผู้วิจัยจึงเล็งเห็นความสำคัญในการสร้างโมเดลให้มีประสิทธิภาพ ที่สามารถตรวจสอบปริมาณอินทรีย์วัตถุ (Organic Matter) ที่เป็นตัวบ่งชี้คุณภาพของดินได้ โดยใช้เทคนิคต่างๆ

ที่กล่าวมา ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลคุณสมบัติทางเคมีของดินที่มีความสัมพันธ์กับปริมาณอินทรีย์วัตถุในดิน ที่เก็บตัวอย่างมาจากสวนผลไม้ในภาคตะวันตก ประกอบด้วยพื้นที่ 3 จังหวัดคือ นครปฐม สมุทรสาคร และสมุทรสงคราม โดยขั้นตอนแรก ใช้วิธีการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) เพื่อลดปัจจัยนำเข้าและขจัด ความสัมพันธ์ (Collinearity) ระหว่างตัวแปรอิสระองค์ประกอบหลักใหม่ที่ได้จะใช้เป็นตัวแปรในการพยากรณ์อินทรีย์วัตถุต่อไปโดยจะพิจารณา วิธีแรก คือวิธีการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression: MLR) ซึ่งเป็นวิธีที่นิยมกันอย่างกว้างขวางในการพยากรณ์ตัวแปรตามที่มีลักษณะข้อมูลแบบต่อเนื่อง และมีความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตามที่เป็นแบบเชิงเส้น ซึ่งข้อมูลที่ได้บางครั้งอาจมีสหสัมพันธ์ระหว่างตัวแปรอิสระและความไม่เป็นเชิงเส้นของอินทรีย์วัตถุ ดังนั้นจึงนำเสนอวิธีที่สองคือวิธีการสร้างโมเดลด้วยวิธีโครงข่ายประสาทเทียม งานวิจัยนี้ใช้วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า (Feed-forward Artificial Neuron Network: FANN) เนื่องจากเป็นวิธีใช้กันอย่างแพร่หลายโดยเฉพาะเป็นวิธีหลักที่ใช้สำหรับการสร้างโมเดลและการพยากรณ์ ซึ่งโมเดลนี้สามารถใช้ได้กับข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น และสามารถแก้ปัญหาที่ยุ่งยากซับซ้อนได้ (Yang et al. 2006 : 213-219) แต่วิธีโครงข่ายประสาทเทียมมีความไวต่อชุดข้อมูลทดสอบมาก และวิธีที่สามคือวิธีจีเนติกอัลกอริทึม (Genetic Algorithm: GA) ซึ่งเป็นวิธีใหม่ที่เห็นกระบวนการชัดเจน รวมทั้งมีความแกร่งต่อชุดข้อมูลทดสอบ ในการทดสอบคำตอบที่ต้องการหาจะกำหนดในรูปแบบโครโมโซม โดยขั้นตอนของจีเนติกอัลกอริทึมจะสร้างและปรับปรุงคุณภาพของประชากรโดยผ่านกระบวนการต่างๆ เพื่อหาคำตอบที่ดีที่สุด แล้วนำผลลัพธ์ที่ได้จากทั้งสามโมเดลมาเปรียบเทียบกันโดยใช้ ค่าวัดประสิทธิภาพ (Performance Indexes) เพื่อให้ได้โมเดลที่มีประสิทธิภาพที่ดีที่สุดในการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน และเป็นการประยุกต์ใช้เทคโนโลยีใหม่ๆ กับข้อมูลทางด้านการเกษตรต่อไป

วัตถุประสงค์ของการศึกษา

1. ศึกษาความสัมพันธ์ของตัวแปรอิสระและลดมิติของข้อมูลโดยใช้การวิเคราะห์องค์ประกอบหลัก

2. เพื่อเปรียบเทียบประสิทธิภาพของโมเดล โดยใช้ตัวแปรนำเข้าที่ได้จากองค์ประกอบหลัก โดยพิจารณาจากโมเดลที่ได้จาก

2.1 วิธีการถดถอยเชิงเส้นพหุ

2.2 วิธีโครงข่ายประสาทเทียมที่มีการเชื่อมโยงแบบไปข้างหน้า

2.3 วิธีจีเนติกอัลกอริทึม

สมมติฐานของการศึกษา

โมเดลในการพยากรณ์ปริมาณอินทรีย์วัตถุในดินได้มีประสิทธิภาพดีที่สุด คือโมเดลที่ได้จากวิธีจีเนติกอัลกอริทึม

ขอบเขตของการศึกษา

1. ข้อมูลดินที่ใช้ในการศึกษาเป็นดินจากพื้นที่เกษตรกรรม ประเภทสวนผลไม้ในภูมิภาคตะวันตก ประกอบด้วย 3 จังหวัดได้แก่ จังหวัดนครปฐม จังหวัดสมุทรสาครและจังหวัดสมุทรสงคราม ซึ่งนำมาวิเคราะห์คุณสมบัติทางเคมีของดิน เก็บข้อมูลตัวอย่างดินมา 58 ชุด จากโครงการ “บทบาทของกรดฮิวมิกและกรดฟัลวิกในดินของพื้นที่เกษตรกรรมต่อการดูดซับจุลธาตุอาหารในดินและโลหะที่เป็นพิษ : กรณีศึกษาภูมิภาคตะวันตกของประเทศไทย”

2. โมเดลที่ใช้พยากรณ์ปริมาณอินทรีย์วัตถุ ได้แก่โมเดลของวิธีการถดถอยเชิงพหุคูณ (MLR) แบบเป็นขั้นตอน (Stepwise multiple regression) วิธีโครงข่ายประสาทเทียม โดยใช้สถาปัตยกรรมโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า (FANN) และวิธีจีเนติกอัลกอริทึม (GA) โดยจะใช้วิธีจีเนติกอัลกอริทึมในการเลือกสัมประสิทธิ์การถดถอยที่ดีที่สุด ให้กับสมการถดถอยเชิงเส้นพหุคูณ

ขั้นตอนการศึกษา

ในงานวิจัยนี้สามารถแบ่งขั้นตอนในการศึกษาได้เป็น 4 ขั้นตอนดังต่อไปนี้

1. การเก็บรวบรวมข้อมูลจากเอกสารและแหล่งข้อมูลที่เกี่ยวข้องกับปัญหาที่ทำการศึกษา ทฤษฎีทางสถิติต่างๆ ที่ใช้ในสร้างโมเดลและกำหนดค่าวัดประสิทธิภาพ

2. วิเคราะห์หาตัวแปรที่เป็นองค์ประกอบหลัก จากตัวแปรคุณสมบัติทางเคมีของดิน ที่จะใช้เป็นตัวแปรพยากรณ์ในโมเดล ด้วยวิธีวิเคราะห์องค์ประกอบ เพื่อลดจำนวนตัวแปรพยากรณ์และขจัดความสัมพันธ์ระหว่างตัวแปร

3. สร้างโมเดลการพยากรณ์ปริมาณอินทรีย์วัตถุในดินด้วยชุดข้อมูลเรียนรู้ แล้วทดสอบโมเดลด้วยชุดข้อมูลทดสอบ เปรียบเทียบประสิทธิภาพของแต่ละโมเดลในการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน

4. สรุปผลที่ได้จากการวิเคราะห์และรวบรวมข้อเสนอแนะ

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้โมเดลที่เหมาะสมในการพยากรณ์ปริมาณอินทรีย์วัตถุ และเป็นเครื่องมือในการตรวจสอบปริมาณอินทรีย์วัตถุในดินต่อไป

2. ได้เรียนรู้หลักการ แนวคิด เกี่ยวกับโครงข่ายประสาทเทียม และวิธีจีเนติกอัลกอริทึม เพื่อประยุกต์ใช้งานในด้านอื่นๆ

3. สามารถนำโมเดลและแนวคิดที่ได้มาประยุกต์ใช้ในการพัฒนางานด้านอื่นๆ ที่มีการพยากรณ์โดยใช้เทคนิคโครงข่ายประสาทเทียม และวิธีจีเนติกอัลกอริทึม เพื่อการพยากรณ์ที่ถูกต้องและมีประสิทธิภาพมากขึ้น

นิยามศัพท์เฉพาะ

1. อินทรีย์วัตถุ (Organic Matter) หมายถึง องค์ประกอบทั้งหลายที่เป็นอินทรีย์สารในดิน รวมทั้งเนื้อเยื่อของพืชและสัตว์ที่ยังไม่เน่าเปื่อย ผลผลิตบางส่วนที่ผ่านการเน่าเปื่อยแล้ว และชีวะมวลของดิน อินทรีย์วัตถุมีผลต่อการเจริญเติบโตของพืชโดยอาศัยคุณสมบัติทางฟิสิกส์, เคมี และชีววิทยาของอินทรีย์วัตถุของดินนั้น มี 3 หน้าที่ดังนี้

1.1 ด้านอาหารพืชเป็นแหล่ง ไนโตรเจน ฟอสฟอรัส สำหรับการเจริญเติบโตของพืช

1.2 ด้านชีววิทยา มีผลอย่างมากต่อการทำงานของจุลินทรีย์ที่ดำรงชีพกับพืชและสัตว์

1.3 ด้านฟิสิกส์และเคมี-ฟิสิกส์ คือ การช่วยเสริมโครงสร้างดินให้ดีขึ้นจึงเป็นการปรับปรุงการไถพรวนให้เกิด ประโยชน์มากขึ้น เพิ่มอากาศ กักเก็บความชื้นและเพิ่มประจุความสามารถของดินในการแลกเปลี่ยน เกลือ แร่ธาตุ และการปรับสมดุลกรด-ด่าง

2. ค่าวัดประสิทธิภาพ (Performance indexes) หมายถึง ดัชนีต่างๆ ที่ใช้ในการวัดหรือประเมินคุณภาพของโมเดล

2.1 ดัชนีการยอมรับ (Index of Agreement : IA) เป็นค่าที่แสดงความถูกต้องของโมเดล ค่าวัดที่ดีที่สุดของค่าดัชนีการยอมรับ มีค่าเท่ากับ 1

$$I.A. = 1 - \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{n}(\left|\hat{Y}_i - \overline{Y}\right| + \left|Y_i - \overline{Y}\right|)^2} \quad (1)$$

2.2 ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) เป็นค่าที่แสดงค่าความผิดพลาดของโมเดล ค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ซึ่งถ้าค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสองให้ค่าวัดที่น้อยกว่า แสดงถึงความแม่นยำที่มากกว่า

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2} \quad (2)$$

2.3 ค่าความคลาดเคลื่อนเอนเอียงเฉลี่ย (Mean Bias Error: MBE) ค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ซึ่งถ้าค่าความคลาดเคลื่อนเอนเอียงเฉลี่ยให้ค่าวัดที่น้อยกว่า แสดงถึงความแม่นยำที่มากกว่า

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i) \quad (3)$$

2.4 ค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) ค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ซึ่งถ้าค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย ให้ค่าวัดที่น้อยกว่า แสดงถึงความแม่นยำที่มากกว่า

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Y}_i - Y_i\right| \quad (4)$$

โดยที่ Y_i คือค่าจริงของปริมาณอินทรีย์วัตถุในดิน

$\hat{Y}_i$ คือค่าพยากรณ์ของปริมาณอินทรีย์วัตถุในดิน

$\overline{Y}$ คือค่าเฉลี่ยของปริมาณอินทรีย์วัตถุในดิน

n คือจำนวนตัวอย่างที่นำมาศึกษา

บทที่ 2
ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

ในงานวิจัยนี้ ผู้วิจัยต้องการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน โดยใช้วิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียม และวิธีจีเนติกอัลกอริทึม โดยศึกษากรณีพื้นที่เกษตรกรรมสวนผลไม้ในภาคตะวันตกของประเทศไทย ในบทนี้จะนำเสนอ ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง เรียงตามลำดับดังนี้

1. การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis)
2. การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression: MLR)
3. วิธีโครงข่ายประสาทเทียม (Artificial Neural Network)
4. วิธีจีเนติกอัลกอริทึม (Genetic Algorithm)
5. งานวิจัยที่เกี่ยวข้อง

การวิเคราะห์องค์ประกอบหลัก

การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) (กัลยา วานิชบัญชา 2551: 182-191) เป็นเทคนิคการลดจำนวนตัวแปร โดยการสร้างเซตของตัวแปรใหม่ให้เป็นฟังก์ชันเชิงเส้นของตัวแปรเดิม และเซตของตัวแปรใหม่จะมีรายละเอียดหรือข้อมูลของตัวแปรเดิม จำนวนตัวแปรใหม่ต้องไม่เกินจำนวนตัวแปรเดิม นั่นคือ กรณีมีตัวแปรเดิม p ตัว จำนวนตัวแปรใหม่จะมี m ตัว จะได้ว่า $m \leq p$

หลักการของการวิเคราะห์องค์ประกอบหลัก

ดังที่ได้กล่าวมาแล้วว่าการวิเคราะห์องค์ประกอบหลักเป็นเทคนิคการลดจำนวนตัวแปรที่มีจำนวนมากเช่น p ตัว โดยการสร้างตัวแปรใหม่ซึ่งเป็นฟังก์ชันของตัวแปรเดิม p ตัว และจะต้องสกัดหรือดึงรายละเอียดหรือค่าแปรปรวนจากตัวแปรเดิมมาไว้ในตัวแปรใหม่ให้มากที่สุด ในที่นี้จะใช้ PC_i แทนองค์ประกอบหลักที่ i ; $i=1, 2, ..., p$ การสร้าง PC_i สรุปได้ดังนี้

ถ้าเวคเตอร์สุ่ม $\boldsymbol{x}' = [X_1, X_2, ..., X_p]$ ซึ่งมีเมตริกซ์ความแปรปรวนร่วมเป็น $\boldsymbol{\Sigma}$ และมีค่าไอเกน (eigenvalue) $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$ โดยเขียนองค์ประกอบหลักให้อยู่ในรูปฟังก์ชันเชิงเส้นของตัวแปร $X_1, X_2, ..., X_p$ แสดงได้ในสมการที่ 1

$$
\begin{aligned}
PC_1 &= \boldsymbol{w}_1'\boldsymbol{x} = w_{11}X_1 + w_{21}X_2 + ... + w_{1p}X_p \\
PC_2 &= \boldsymbol{w}_2'\boldsymbol{x} = w_{21}X_1 + w_{22}X_2 + ... + w_{2p}X_p \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
PC_k &= \boldsymbol{w}'_p\boldsymbol{x} = w_{21}X_1 + w_{22}X_2 + ... + w_{2p}X_p
\end{aligned}
$$

ดังนั้น ความแปรปรวนคือ $Var(PC_i) = \boldsymbol{w}'_i \boldsymbol{\Sigma}\, \boldsymbol{w}_i \ , i = 1, 2, ..., p$

ความแปรปรวนร่วม $Cov(PC_i, PC_k) = \boldsymbol{w}'_i \boldsymbol{\Sigma}\, \boldsymbol{w}_i \ ; i \neq k, i, k = 1, 2, ..., p$

โดยที่ $PC_1, PC_2, ..., PC_k$ เป็นตัวแปรใหม่หรือตัวประกอบหลัก p ตัว และ $\boldsymbol{w}_{ij}$ เป็นสัมประสิทธิ์น้ำหนักของตัวแปรเดิม $\boldsymbol{X}_{ij}$ ที่มีต่อตัวประกอบหลัก PC_i โดยที่ $PC_1, PC_2, ..., PC_k$ ไม่มีความสัมพันธ์กัน

PC_1 เป็นตัวแปรใหม่ตัวที่ 1 หรือองค์ประกอบหลักที่ 1 ซึ่งเป็นฟังก์ชันเชิงเส้นของตัวแปรเดิม ซึ่งทำให้ค่าความแปรปรวน Var(PC_1) $= \boldsymbol{w}'_i \boldsymbol{\Sigma}\, \boldsymbol{w}_i$ มีค่ามากที่สุด หรือ

PC_1 = ฟังก์ชันเชิงเส้น $\boldsymbol{w}_1'\boldsymbol{x}$ ที่ทำให้ Var($\boldsymbol{w}_1'\boldsymbol{x}$) มีค่ามากที่สุดซึ่งมี $w_1'w_1 = 1$ หรือ

$\boldsymbol{w}_1'\boldsymbol{w}_1 = w^2_{11} + w^2_{12} + ... + w^2_{1p} = 1$

PC_2 = ฟังก์ชันเชิงเส้นที่ 2 ของ $X_1, X_2, ..., X_p$ = $\boldsymbol{w}_2'\boldsymbol{x}$ ที่ทำให้ Var ($\boldsymbol{w}_2'\boldsymbol{x}$) มีค่ามากที่สุดโดยมีเงื่อนไขว่า $\boldsymbol{w}_2'\boldsymbol{w}_2$ = 1 และ $Cov\ (\boldsymbol{w}_1'\boldsymbol{x}, \boldsymbol{w}_2'\boldsymbol{x}) = Cov(PC_1, PC_2) = 0$ หรือ PC_1 ไม่มีความสัมพันธ์กับ PC_2

PC_k = ฟังก์ชันเชิงเส้นที่ k ของ $X_1, X_2, ..., X_p$ = $\boldsymbol{w}'_k\boldsymbol{x}$ ที่ทำให้ Var($\boldsymbol{w}'_k\boldsymbol{x}$) มีค่ามากที่สุด โดยมีเงื่อนไขว่า $\boldsymbol{w}'_k\boldsymbol{w}_k = 1$ และ $Cov\ (\boldsymbol{w}'_k\boldsymbol{x}, \boldsymbol{w}'_i\boldsymbol{x}) = 0$ เมื่อ $i < k$ หรือ PC_k ไม่มีความสัมพันธ์กับ PC_i โดยที่ $i < k$

ความสัมพันธ์ระหว่างตัวแปรเดิมกับตัวประกอบหลัก

ตัวแปรใหม่หรือองค์ประกอบหลักจะไม่มีความสัมพันธ์กันเอง ดังนั้นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง PC_i และ PC_j ($i \neq j$) จะเป็นศูนย์ แต่องค์ประกอบหลักกับตัวแปรเดิมจะมี

ความสัมพันธ์โดยวัดความสัมพันธ์ระหว่าง PC_i และ $\boldsymbol{X}_j$ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์และเรียกค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างองค์ประกอบหลักและตัวแปรเดิมว่าค่าน้ำหนัก (loading) ดังนั้นค่า น้ำหนักจึงเป็นค่าที่แสดงถึงอิทธิพลของตัวแปรเดิมที่มีต่อการสร้างตัวประกอบหลัก ถ้าค่าน้ำหนัก ของตัวแปรเดิมใดมีค่ามาก (ใกล้ +1 หรือใกล้ -1) แสดงว่าตัวแปรเดิมนั้นมีความสำคัญหรือมีส่วนร่วมในการสร้างตัวประกอบหลักมาก แสดงความสัมพันธ์ระหว่างตัวแปรเดิมที่ปรับให้อยู่ในรูปมาตรฐานกับตัวประกอบหลักดังนี้

$$
\begin{aligned}
Z_1 &= l_{11}PC_1 + l_{11}PC_2 + \ldots + l_{1p}PC_p \\
Z_2 &= l_{22}PC_1 + l_{22}PC_2 + \ldots + l_{2p}PC_p \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
Z_p &= l_{p1}PC_2 + l_{p2}PC_2 + \ldots + l_{pp}PC_p
\end{aligned}
$$

โดยที่ $\boldsymbol{Z}_j$ เป็นตัวแปร $\boldsymbol{X}_j$ ที่ปรับให้อยู่ในรูปมาตรฐาน; $j = 1, 2, \ldots, p$

ความสามารถของ PCA บนชุดข้อมูลที่ใช้ในการศึกษานี้ถูกวัดโดยใช้การทดสอบ Bartlett's sphericity โดย Bartlett ได้พัฒนาการทดสอบที่เรียกว่า Sphericity test เพื่อตรวจสอบว่าเมทริกซ์สัมประสิทธิ์สหสัมพันธ์ของตัวแปร p ตัว สามารถแบ่งเป็นปัจจัยร่วมหรือกลุ่มที่ตัวแปรมีส่วนร่วมกันหรือสัมพันธ์กันได้หรือไม่ โดยที่ Bartlett's sphericity จะมีการแจกแจงโดยประมาณแบบไคกำลังสอง (χ^2) ซึ่งเป็นฟังก์ชันของค่าดีเทอร์มิแนนต์ของเมทริกซ์สัมประสิทธิ์สหสัมพันธ์ ($\boldsymbol{R}$) แสดงได้ในสมการที่ (5)

$$
\chi^2 = -\left[(n-1) - \frac{2(p+5)}{6}\right] \ln|\boldsymbol{R}| \qquad (5)
$$

โดยที่ χ^2 มีองศาอิสระเท่ากับ $\frac{1}{2}p(p-1)$

$\ln|\boldsymbol{R}|$ คือค่า *log* ของดีเทอร์มิแนนต์ของเมตริกซ์สหสัมพันธ์ $\boldsymbol{R}$

p คือจำนวนตัวแปร

n คือจำนวนข้อมูล

การหาค่าดีเทอร์มิแนนต์ของเมตริกซ์สหสัมพันธ์ $\boldsymbol{R}$ สามารถใช้ค่าไอเกนได้ดังแสดงในสมการที่ 6

$$|\boldsymbol{R}| = \prod_{i=1}^{p} \lambda_i \qquad (6)$$

โดยที่ λ_i คือค่าไอเกนของตัวแปรที่ i ; $i= 1,2,...,p$

สมมติฐานว่างที่พิจารณาคือตัวแปรทุกตัวไม่มีสหสัมพันธ์ต่อกัน (เมทริกซ์สหสัมพันธ์เป็นเมทริกซ์เอกลักษณ์ ($\boldsymbol{R} = \boldsymbol{I}$) ถ้าตัวแปรไม่มีความสัมพันธ์กัน จะได้ว่า $\boldsymbol{R} = \boldsymbol{I}$ ซึ่งจะทำให้ค่าไอเกนของตัวแปรมีค่าเข้าใกล้ 1 หรือ $|\boldsymbol{R}|$ มีค่าใกล้ -1 ซึ่งมีผลทำให้ $\ln|\boldsymbol{R}|$ ลู่เข้าสู่ศูนย์ ค่า χ^2 จะต่ำจึงไม่สามารถปฏิเสธ H_0 ได้ นั่นคือไม่สมควรใช้การวิเคราะห์องค์ประกอบร่วมกับตัวแปรดังกล่าว

แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่ามาก หรือตัวแปรมีความสัมพันธ์กัน จะทำให้ค่าไอเกนของบางตัวแปรมีค่ามากกว่าหนึ่ง ขณะที่ค่าไอเกนของบางตัวแปรมีค่าใกล้ศูนย์ จะทำให้ $\ln|\boldsymbol{R}|$ ติดลบมาก ซึ่งมีผลให้ χ^2 สูง จึงปฏิเสธ H_0 ได้จึงสมควรใช้การวิเคราะห์องค์ประกอบได้

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Regression Analysis) ถ้ามีตัวแปรอิสระ k ตัว ($X_1, X_2,..., X_k$) ที่มีความสัมพันธ์กับตัวแปรตาม Y โดยที่ความสัมพันธ์อยู่ในรูปเชิงเส้น จะได้สมการความถดถอยเชิงพหุ ซึ่งแสดงความสัมพันธ์ระหว่าง Y และ $X_1, X_2,..., X_k$ ดังสมการที่ (7)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + e \qquad (7)$$

โดยที่ Y คือตัวแปรตาม (Dependent Variable) เนื่องจากค่าของ Y ขึ้นอยู่กับค่าของ X

X คือตัวแปรอิสระ (Independent Variable)

β_0 คือส่วนตัดแกน Y หรือค่าของ Y เมื่อกำหนดให้ $X_1=X_2=...=X_k=0$

$\beta_1, \beta_2, ..., \beta_k$ คือ สัมประสิทธิ์ความถดถอยเชิงส่วน (Partial Regression Coefficient)

โดยที่ β_i เป็นค่าที่แสดงถึงการเปลี่ยนแปลงของ Y เมื่อ X_i เปลี่ยนไป 1 หน่วย โดยที่ตัวแปรอิสระ X ตัวอื่นๆ มีค่าคงที่ จากการประมาณค่า β_i ด้วย b_i และประมาณค่า β_0 ด้วย a จะทำให้สมการถดถอยเชิงพหุเป็นดังสมการที่ (8)

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + ... + b_k X_k \qquad (8)$$

โดยที่ $\hat{Y}$ คือค่าประมาณ หรือค่าพยากรณ์ของตัวแปรของ Y

$e = Y - \hat{Y}$ คือค่าคลาดเคลื่อนหรือความแตกต่างระหว่างค่าจริงกับค่าพยากรณ์ การประมาณค่า β_i ด้วย b_i และประมาณค่า β_0 ด้วย a จะใช้วิธีกำลังสองน้อยที่สุด (Least Square Method) ซึ่งเป็นการหาค่า $a, b_1, b_2,..., b_k$ ที่ทำให้

$$\sum_{i=1}^{n} e_i^2 = \sum (Y_i - \hat{Y})^2 \text{ มีค่าน้อยที่สุด}$$

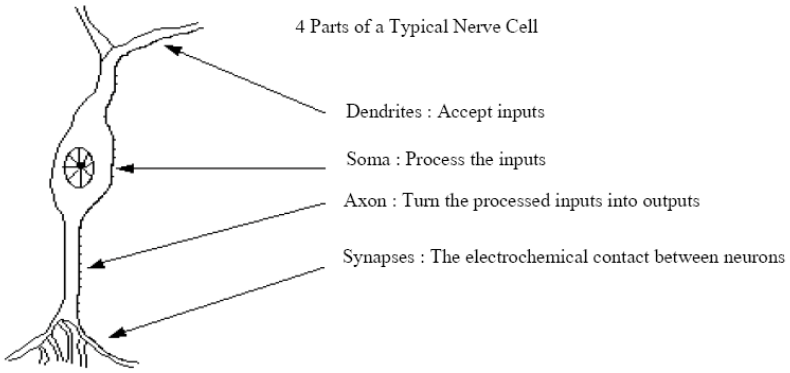
วิธีโครงข่ายประสาทเทียม

แบบจำลองโครงข่ายประสาทเทียม (Artificial neural network) (ศรัลย์ ปานศรีพงษ์ 2550 : 13-22) ประกอบด้วย ระบบการคำนวณแบบไม่เป็นเชิงเส้น ซึ่งมีวิธีการดำเนินระบบเป็นแบบขนาน และมีรูปแบบการเรียนรู้แบบโครงข่ายใยประสาทชีวภาพ (Lippmann 1987 : 4-22) โดยประกอบไปด้วยนิวรอน (โหนดหรือหน่วยประมวลผล) ซึ่งรวมกันอยู่เป็นชั้น ๆ ซึ่งสามารถรับข้อมูลเข้าได้หลายค่า และสามารถคำนวณผล โดยจะให้ผลลัพธ์ค่าเดียวหรือหลายค่าก็ได้ (Klimasauskas 1993 : 47-72) ซึ่งการคำนวณในระบบประกอบไปด้วยฟังก์ชันง่าย ๆ เช่นฟังก์ชันการรวม และ ฟังก์ชันการคูณ โดยมีความสามารถในการเรียนรู้จากตัวอย่างหลาย ๆ ตัวอย่าง ซึ่งจะหาแนวทางการแก้ปัญหา แม้แต่ข้อมูลที่ถูกป้อนเข้ามาไม่สมบูรณ์หรือผิดพลาด ระบบจะเปรียบเทียบผลลัพธ์ที่คลาดเคลื่อน และปรับเปลี่ยนวิธีการประมวลผลเพื่อให้ผลลัพธ์ถูกต้องที่สุด ระบบจะประมวลผลข้อมูลโดยคอมพิวเตอร์อย่างรวดเร็ว (Flood and Kartam 1994 : 149-162)

แนวคิดพื้นฐานของโครงข่ายประสาทเทียม

Lippmann (Lippmann 1987) และ Kireetoh (Kireetoh 1995) ได้อธิบายถึงนิวรอนว่าคล้ายกับระบบประสาทของมนุษย์ โดยรับสัญญาณข้อมูลที่ส่งเข้ามา และจะถูกกระตุ้น แต่ละเซลล์ประกอบด้วย ปลายในการรับกระแสประสาท เรียกว่า "เดนไดรท์" (Dendrite) ซึ่งเป็นข้อมูลป้อนเข้า

และปลายในการส่งกระแสประสาท เรียกว่า "แอคซอน" (Axon) แสดงดังภาพที่ 1 ซึ่งเป็นเหมือนข้อมูลที่ส่งออกมาของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอก หรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียส ซึ่งจะเป็นตัวตัดสินว่า ต้องกระตุ้นเซลอื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลอื่น ๆ ต่อไปผ่านทางแอคซอนของมัน ผลการกระตุ้นด้วยสิ่งเร้าที่เหมือนหรือมีลักษณะพิเศษบางอย่างเหมือนกัน จะให้ผลลัพธ์สุดท้ายเป็นค่าที่ค่อนข้างแน่นอน



ภาพที่ 1 โครงสร้างระบบประสาทในสมอง

ที่มา : ศรัลย์ ปานศรีพงษ์, “การพัฒนาแบบจำลองโครงข่ายใยประสาทประดิษฐ์และแบบจำลองความถดถอยเชิงพหุ เพื่อทำนายความเข้มข้นของ PM10 ในพื้นที่กรุงเทพมหานครและเทศบาลนครราชสีมา” (วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชา วิศวกรรมสิ่งแวดล้อม บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีสุรนารี, 2550), 14.

โครงข่ายประสาทเทียมเป็นปัญญาประดิษฐ์ (Artificial Intelligence) ที่สร้างเลียนแบบระบบประสาทของมนุษย์ (Lippmann 1987:4-22, Klimasauskas 1993: 47-72, Medsker et al. 1993: 3-26) โดยระบบการทำงานต่าง ๆ ของโครงข่ายประสาทเทียม มีลักษณะคล้ายกับระบบประสาทของมนุษย์ คือ มีการเรียนรู้จากประสบการณ์ที่ได้รับ และสามารถ ให้คำตอบได้แม้แต่ข้อมูลที่ป้อนเข้าจะผิดพลาดหรือไม่สมบูรณ์ โดยหาวิธีการแก้ปัญหาจากประสบการณ์การเรียนรู้ที่ผ่านมา และสามารถที่จะพัฒนาให้ไปเป็นโปรแกรมที่สามารถโต้ตอบกับมนุษย์ได้

การสร้างโมเดลโครงข่ายประสาทเทียม

โครงข่ายใยประสาทเทียม คือ เครื่องจักรการเรียนรู้ ตั้งอยู่บนพื้นฐานความคิดเกี่ยวกับการปรับปรุงตัวแปรควบคุมต่าง ๆ ภายในด้วยตัวมันเอง โดยมีองค์ประกอบภายในระบบซึ่งประกอบไปด้วย 5 องค์ประกอบหลัก คือ หน่วยการเรียนรู้, โครงข่ายใยประสาท, แผนการเรียนรู้, กระบวนการเรียนรู้ และกระบวนการวิเคราะห์ (Adeli 1992: 1-10) ซึ่ง Elazouni et al. (1997) ได้จำแนกส่วนประกอบของโครงข่ายใยประสาทประดิษฐ์ เป็น 3 ขั้น คือ 1) การออกแบบ 2) การสร้างแบบจำลอง และ 3) การทดสอบและหาผลลัพธ์ โดยขั้นการออกแบบจะประกอบไปด้วยกัน 2 ส่วน คือ การวิเคราะห์โครงสร้างของปัญหา และการวิเคราะห์ปัญหา ส่วนขั้นการสร้างแบบจำลอง จะแบ่งย่อยออกเป็น 3 ขั้นตอนคือ 1) การเลือกข้อมูล 2) การเลือกรูปแบบโครงข่าย 3) การสอนและการทดสอบโครงข่าย

ข้อมูลป้อนเข้า

โครงข่ายใยประสาทประดิษฐ์ ประกอบไปด้วยตัวแปรอิสระ หรือ ข้อมูลป้อนเข้าและตัวแปรตาม หรือผลลัพธ์ โดยหลักการเลือกตัวแปรที่ใช้ในโครงข่ายที่เกี่ยวข้องจะมี 2 แบบ (Smith, 1993) คือ วิธีแรก ข้อมูลจะต้องถูกแปลงรูปให้อยู่ในรูปที่เหมาะสม และวิธีที่สอง คือการเลือกข้อมูลโดยใช้พื้นฐานระหว่าง predictiveness และ covariance โดยปกติแล้ว ตัวแปรอิสระที่ถูกเลือกจะมีความสามารถในการทำนายผลหากตัวแปรที่เลือกมีความสัมพันธ์กันในทางตรงกันข้ามหากตัวแปรอิสระ 2 ตัวมีความสัมพันธ์ต่อกัน จะทำให้แบบจำลองมีความอ่อนไหว (sensitive) และเกิดปัญหาที่เรียกว่า over fitting และ limit generalization ด้วยเหตุผลนี้ การเลือกข้อมูลจะต้องเลือกเฉพาะตัวแปรอิสระที่มีความสามารถทำนายผลเพื่อให้ได้ ผลลัพธ์หรือตัวแปรตาม โดยตัวแปรอิสระที่เลือกมานั้นจะต้องไม่มีความสัมพันธ์กัน แต่อย่างไรก็ดี ก็ขึ้นอยู่กับรูปแบบของโครงข่ายที่ใช้และเพื่อที่จะลดจำนวนตัวอย่างที่ใช้ในการสอนและเวลาที่ใช้ในการเรียนรู้ ก็ควรจะต้องมีการคัดเลือกข้อมูลป้อนเข้าให้เหมาะสม เพราะการคัดเลือกข้อมูลเป็นปัจจัยที่สำคัญในการสร้างแบบจำลอง (Wu and Lim 1993: 61-66)

ชั้นซ่อน (Hidden Layer)

ชั้นซ่อนเป็นชั้นประมวลผลที่อยู่ระหว่างชั้นข้อมูลป้อนเข้า และชั้นแสดงผลลัพธ์โดยปกติแล้ว ชั้นซ่อนอาจมีมากกว่า 1 ชั้น โดยโครงข่ายจะสามารถประมวลหาฟังก์ชั่นที่เหมาะสม จากปัญหาที่

ซับซ้อนได้หากมีชั้นซ่อนที่มากพอ (Lippmann 1987: 4-22) ข้อมูลที่ได้จากชั้นซ่อนจะได้เป็นตัวแปรใหม่ที่จะถูกส่งต่อให้กับชั้นแสดงผลลัพธ์ หรือชั้นตัวแปรตาม ถ้าโครงข่ายแบบแพร่กลับ (Back-propagation) มีชั้นซ่อนที่น้อยเกินไปแล้วจะทำให้โครงข่ายไม่สามารถที่จะหาทางแก้ปัญหาได้ (Karunasekera, 1992) แต่ถึงอย่างไรก็ตามถ้าเกิดโครงข่ายมีชั้นซ่อนที่มากจนเกินไป จะทำให้โครงข่ายมีระยะเวลาในการเรียนรู้นาน หากมีชั้นซ่อนที่มากก็จะไม่ช่วยให้โครงข่ายมีประสิทธิภาพมากขึ้น หรืออีกนัยหนึ่ง การที่มีโหนดในแต่ละชั้นที่มากเกิน จะทำให้โครงข่ายไม่สามารถที่จะหาจุดสิ้นสุดได้ การที่มีโหนดในชั้นซ่อนมากเกินจะทำให้เกิดปัญหาที่เรียกว่า over fitting โดยโครงข่ายจะจำลองโครงสร้างใหม่เกินความเป็นจริงจาก noise ของข้อมูล แทนที่จะหาฟังก์ชั่นที่เหมาะสมในการวิเคราะห์ปัญหาให้ถูกต้องตามที่ควรจะเป็น (Smith, 1993)

ดังนั้น การที่จะทำให้โครงข่ายเกิดประสิทธิภาพสูงสุด ต้องกำหนดให้มีโหนดในชั้นซ่อนอยู่ให้น้อยเท่าที่จะเป็นไปได้ (Khan, Topping and Bahreininejad 1993: 81-94) Berke และ Hajela (1991) ได้ให้ความเห็นว่า จำนวนของโหนดในชั้นซ่อนควรอยู่ระหว่างค่าเฉลี่ยของผลรวมของโหนดในชั้นข้อมูลป้อนเข้า และชั้นแสดงผลลัพธ์ Soemardi (1996) ได้แสดงความเห็นว่า จำนวนโหนดในชั้นซ่อนควรจะมีค่าเท่ากับร้อยละ75 ของโหนดในชั้นข้อมูลป้อนเข้า ดังนั้นจึงสรุปจากข้อคิดเห็นได้ว่า จำนวนโหนดในชั้นซ่อนที่มากที่สุดควรจะเท่ากับผลรวมของโหนดในชั้นข้อมูลป้อนเข้า และชั้นแสดงผลลัพธ์ และจำนวนโหนดที่น้อยที่สุดควรจะเท่ากับร้อยละ 75 ของโหนดชั้นข้อมูลป้อนเข้า หรือเท่ากับค่าเฉลี่ยของผลรวมของโหนดในชั้นข้อมูลป้อนเข้าและชั้นแสดงผลลัพธ์

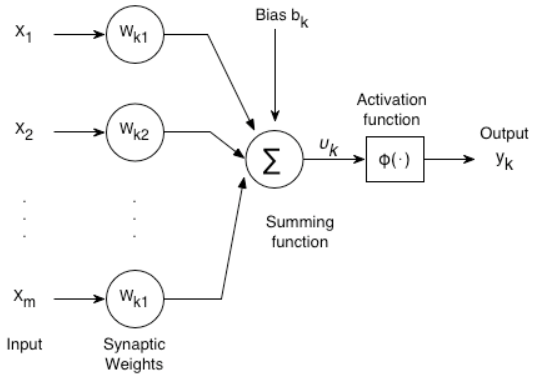
ค่าถ่วงน้ำหนักและไบแอส (Weights and biases)

ค่าถ่วงน้ำหนักถูกแทนด้วยตัวเลขเพื่อแสดงถึงความแรงในการเชื่อมต่อของโหนดแต่ละโหนดที่ถูกเชื่อมต่อเข้าด้วยกัน ซึ่งผลรวมของค่าถ่วงน้ำหนักที่ป้อนเข้าจะไปปรับปรุงการประมวลผลในแต่ละโหนด ค่าถ่วงน้ำหนักคือค่าความแรงสัมพัทธ์ (ในทางคณิตศาสตร์) ของการเชื่อมต่อซึ่งส่งผลต่อการส่งผ่านข้อมูลจากชั้นหนึ่งไปยังชั้นต่อไป (Medsker et al. 1993: 3-26) โดยปกติค่าถ่วงน้ำหนักจะถูกกำหนด และเริ่มป้อนเข้าสู่โครงข่ายในขั้นตอนการเรียนรู้ ซึ่งต้องมีหลักการในการกำหนดค่าเพื่อที่จะให้โครงข่ายสามารถแก้โจทย์ปัญหา และลดเวลาการเรียนรู้ได้ สำหรับโครงข่ายใด ๆ ค่าถ่วง

น้ำหนักจะมีค่าเท่ากับผลคูณของจำนวนโหนดของทุก ๆ การเชื่อมต่อ และค่าของไบแอสจะเท่ากับผลรวมของจำนวนโหนดของทุก ๆ การเชื่อมต่อ

ฟังก์ชั่นการรวมและฟังก์ชั่นการแปลงค่า (Summation and transfer function)

ฟังก์ชั่นการรวม คือ ฟังก์ชั่นการหาผลเฉลี่ยค่าถ่วงน้ำหนักของทุก ๆ โหนดที่เชื่อมต่อกัน โดยมีขั้นตอนคือ นำค่าของข้อมูลป้อนเข้า (Input) ในแต่ละโหนดคูณกับค่าถ่วงน้ำหนักของแต่ละโหนดและรวมผลลัพธ์ของทุก ๆ โหนดเข้าด้วยกัน แสดงดังภาพที่ 2 ส่วนฟังก์ชั่นการแปลงค่าคือความสัมพันธ์ระหว่างระดับการกระตุ้นภายในโหนด (N) กับผลลัพธ์ที่ได้ (Output) โดยจะอยู่ในรูปของฟังก์ชั่นซิกมอยด์ f(N) โดยมีข้อแม้ว่า 1) มีความต่อเนื่อง 2) ค่าของฟังก์ชั่นซิกมอยด์จะต้องเพิ่มขึ้นเมื่อ N เพิ่มขึ้น (Smith 1993)



ภาพที่ 2 กระบวนการทำงานของโครงข่ายประสาทเทียม

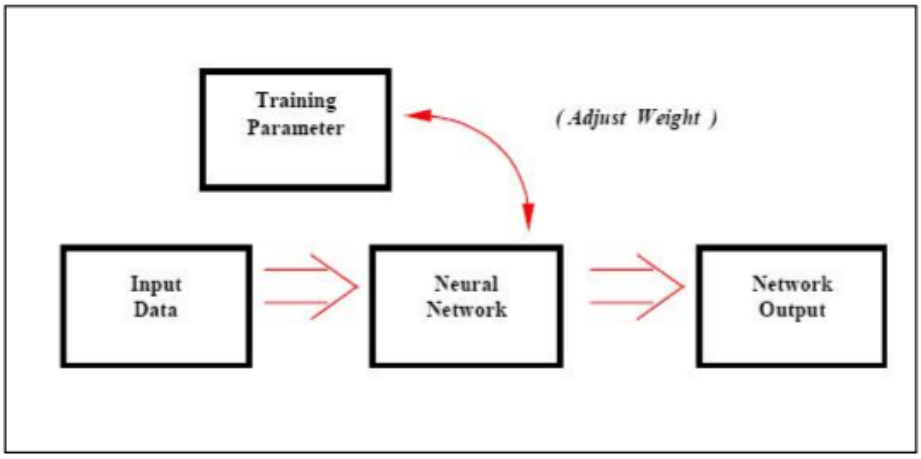
ที่มา : Simon Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed. (New York : Prentice Hall, 1998), 11.

กระบวนการสอนหรือการเรียนรู้ (Training or Learning)

กระบวนการเรียนรู้เป็นกระบวนการหนึ่งในโครงข่ายที่เรียนรู้จากความผิดพลาดโดยมีหลักการ 3 ข้อ คือ 1) คำนวณหาคำตอบ 2) ตรวจสอบคำตอบว่าถูกต้องหรือไม่ 3) ปรับแก้ค่าถ่วงน้ำหนักแล้วคำนวณใหม่อีกครั้ง (Medsker et al 1993: 3-26) กระบวนการสอน คือการนำข้อผิดพลาดจากการคำนวณครั้งก่อนมาปรับแก้ค่าถ่วงน้ำหนักให้กับการสอนในรอบต่อไป ทำให้คำตอบที่ได้มี

ความถูกต้องสูงขึ้น (Klimasauskas 1993: 47-72) ในขั้นตอนการเรียนรู้จะเป็นกระบวนการที่เกี่ยวข้องกับ การปรับค่าน้ำหนักของแต่ละโหนด โดยอาศัยจากประสบการณ์การเรียนรู้ของโครงข่าย การเรียนรู้ของโครงข่ายในแต่ละรอบจะนำผลลัพธ์ที่ประมวลได้ มาเปรียบเทียบกับค่าจริงที่ได้จากการทดลอง ซึ่งอาจจะมีความคลาดเคลื่อน หลังจากนั้นโครงข่ายจะย้อนกลับไปเรียนรู้ในรอบต่อไปพร้อมกับการปรับแก้ค่าถ่วงน้ำหนักเพื่อให้การประมวลผลรอบต่อไปมีความแม่นยำมากขึ้น Bhokha (1998) ได้กล่าวว่าการปรับแก้อาจจะเป็นการปรับขึ้นหรือปรับลงก็ได้ Klimasauskas (1993) กล่าวว่า การวัดผลว่าโครงข่ายสามารถเรียนรู้ได้ดีเพียงใดจะดูจากตัวชี้วัดต่าง ๆ เช่น ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error) ในชั้นแสดงผลลัพธ์ Lippmann (1987) และ Smith (1993) ได้กล่าวไว้ว่ากระบวนการเรียนรู้สามารถแบ่งออกเป็น 2 ลักษณะ คือ

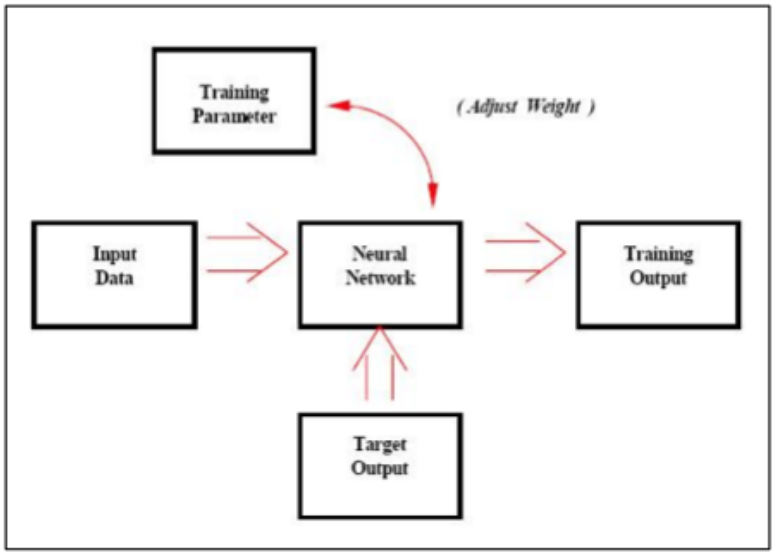
การเรียนรู้แบบไม่มีครูสอน (unsupervised training) ได้คิดค้นโดย Kohonen (1984) ซึ่งแตกต่างจากแบบจำลองที่เลียนแบบระบบสมองของมนุษย์ โดยไม่ต้องใช้ผลลัพธ์จริงมาทำการเปรียบเทียบ แต่จะใช้คุณสมบัติทางสถิติของข้อมูลชุดทดสอบมาจัดกลุ่มเป็นหมวดหมู่หลังจากที่ป้อนข้อมูลเข้าไปแล้ว แบบจำลองจะทำการประมวลผลลัพธ์ที่เป็นไปได้ออกมาเป็นชุด ๆ



ภาพที่ 3 การเรียนรู้แบบไม่มีการเรียนการสอน (Unsupervised Learning)

ที่มา : Artificial Neural Network [Online], accessed 20 November 2009. Available from http://202.28.94.55/web/320417/2548/work1/g26/Files/Report_Neural%20Network.doc

การเรียนรู้แบบมีครูสอน (supervised training) ซึ่งจะประกอบไปด้วย คู่อันดับของข้อมูลป้อนเข้า และผลลัพธ์จริง ซึ่งเมื่อโครงข่ายเริ่มการเรียนรู้จากข้อมูลป้อนเข้า และคำนวณหาผลลัพธ์ได้แล้ว จึงจะนำไปเปรียบเทียบกับผลลัพธ์จริง เพื่อหาความคลาดเคลื่อนซึ่งความคลาดเคลื่อนดังกล่าวจะถูกส่งกลับเข้าไปยังโครงข่ายพร้อมกับการปรับแก้ค่าถ่วงน้ำหนักเพื่อให้โครงข่ายคำนวณผลลัพธ์ใหม่ให้มีความคลาดเคลื่อนน้อยที่สุด



ภาพที่ 4 การเรียนรู้แบบมีการเรียนการสอน (Supervised Learning)

ที่มา : Artificial Neural Network [Online], accessed 20 November 2009. Available from http://202.28.94.55/web/320417/2548/work1/g26/Files/Report_Neural%20Network.doc

ตัวแบบโครงข่ายประสาทเทียมมีหลากหลายรูปแบบด้วยกัน ในงานวิจัยนี้จะใช้ตัวแบบ feed forward back propagation neural network ซึ่งเป็นการเรียนรู้แบบมีการเรียนการสอน โหนดในชั้นข้อมูลเข้าทุกโหนดจะถูกเชื่อมต่อไปข้างหน้ากับทุกโหนดที่อยู่ในชั้นถัดไป และโหนดในชั้นถัดไปจะถูกเชื่อมกับชั้นถัดไปอีกตามลำดับ จนกระทั่งถึงชั้นผลลัพธ์ การเรียนรู้แบบแพร่กลับนี้ เป็นที่ยอมรับกันอย่างแพร่หลายโดยเป็นการเรียนรู้แบบมีครูสอน และมีระบบการเชื่อมโยงแบบเคลื่อนไปข้างหน้าหลายชั้น (multilayer feed forward) (Bhokha 1998) ดังภาพที่ 5 สำหรับวิธีการเรียนรู้แบบแพร่กลับ จะมีการปรับค่าถ่วงน้ำหนัก (w) ในทุก ๆ รอบการเรียนรู้เพื่อให้เกิดค่าความผิดพลาดที่

น้อยที่สุด โดยจะเริ่มต้นปรับแก้ค่าถ่วงน้ำหนักตั้งแต่ชั้น output หลังจากนั้นทำการปรับย้อนกลับมาที่ชั้นซ่อนชั้นสุดท้าย และชั้นซ่อนชั้นต่อ ๆ มา จนกระทั่งถึงชั้นซ่อนชั้นแรก ดังสมการที่ 9 – 13

$$\mathbf{w_{ใหม่}} = \mathbf{w_{เดิม}} + \Delta w \tag{9}$$

โดยที่ การปรับค่าถ่วงน้ำหนักที่ชั้น Output layer มีค่าเท่ากับ

$$\Delta w = \eta y_i (t_k - z_k) f'(net_{jk})_o \tag{10}$$

กำหนดให้ $\delta_k = (t_k - z_k) f'(net_{jk})_o$ ดังนั้น

$$\Delta w = \eta \delta_k y_i \tag{11}$$

โดยที่ การปรับค่าถ่วงน้ำหนักที่ชั้น Hidden layer มีค่าเท่ากับ

$$\Delta w = \eta x_i f'(net_{ij})_h \Sigma_k \delta_k w_{jk} \tag{12}$$

กำหนดให้ $\delta_j = f'(net_{ij})_h \Sigma_k \delta_k w_{jk}$ ดังนั้น

$$\Delta w = \eta \delta_j x_i \tag{13}$$

โดยที่ h = ชั้นซ่อน

o = ชั้นผลลัพธ์

w_{ij} = ค่าถ่วงน้ำหนักชั้นซ่อน

w_{jk} = ค่าถ่วงน้ำหนักชั้นผลลัพธ์

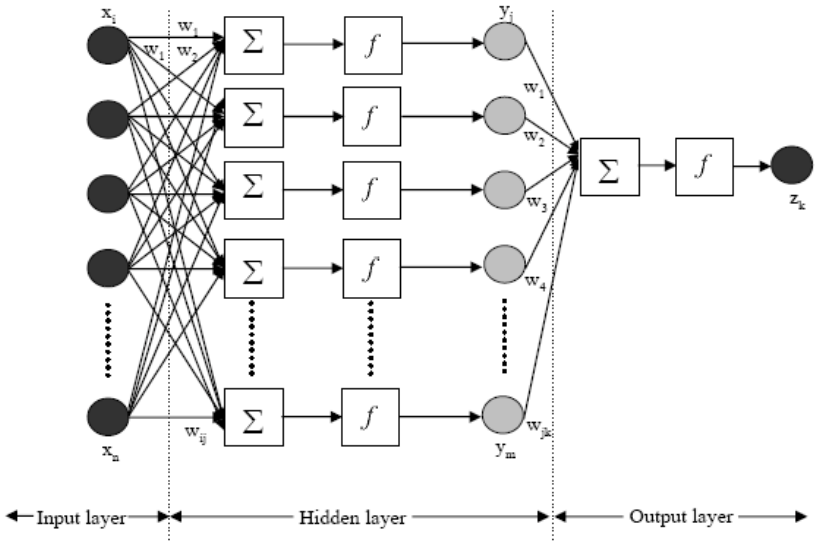
x_i = ข้อมูลป้อนเข้า

y_j = ผลลัพธ์ (ชั้นซ่อน)

z_k = ผลลัพธ์ (ชั้นผลลัพธ์)

t_k = ค่าจริง

η = อัตราการเรียนรู้



ภาพที่ 5 โครงข่ายประสาทเทียมแบบแพร่กลับ

ที่มา : ศรัลย์ ปานศรีพงษ์, “การพัฒนาแบบจำลองโครงข่ายใยประสาทประดิษฐ์และแบบจำลองความถดถอยเชิงพหุ เพื่อทำนายความเข้มข้นของ PM10 ในพื้นที่กรุงเทพมหานครและเทศบาลนครราชสีมา” (วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชา วิศวกรรมสิ่งแวดล้อม บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีสุรนารี, 2550), 20.

โดยในงานวิจัยนี้เลือกใช้ Levenberg –Maquardt back propagation เป็นอัลกอริทึมในการเรียนรู้ที่เหมาะสมกับปัญหาขนาดเล็ก เทคนิคนี้ใช้กับสถาปัตยกรรมของโครงข่ายที่มีจำนวนค่าน้ำหนักน้อยกว่า 100 โดยอัลกอริทึมนี้ใช้ Hessian matrix จำกัดจำนวนค่าน้ำหนัก เพื่อป้องกันการ overfitting ของโครงข่ายรวมทั้งหลีกเลี่ยง bad local minima (Matignon , 2007) โดยวิธีการฝึกเครือข่ายแบบ Levenberg-Maquardt (Adineh และคณะ 2008 : 1000-1007) จะเป็นวิธีการของ Newton ซึ่งใช้อินเวิร์สเมทริกซ์ของ Hessian ในการทำหน้าที่เป็นผลของสัดส่วนการเรียนรู้คูณด้วยทิศทางการลดลงของค่าผิดพลาดแบบกราเดียนส์ ซึ่ง Hessian matrix จะเป็นดัชนีค่าความผิดพลาด (error index) สำหรับอนุพันธ์อันดับที่สองของค่าน้ำหนักและฟังก์ชันควอดราติก (quardratic function) เพื่อให้ได้ค่าจุดถัดไปต่ำสุดตามต้องการ ดังนั้นอัตราการลู่เข้าด้วยวิธีการแบบ Newton จะเร็วกว่าแบบ back propagation ธรรมดา สมการการปรับค่าน้ำหนักด้วยวิธีของ Levenberg-Maquardt แสดงดังสมการที่ 14

$$x^{k+1} = x^{k} + (J^{T}J + \mu I)^{-1} J^{T} e \qquad (14)$$

เมื่อ x^{k+1} คือค่าน้ำหนักที่ถูกคำนวณในขั้นตอนก่อนหน้า

x^{k} คือค่าน้ำหนักที่ได้ในขั้นตอนใหม่

μ คือค่าน้ำหนักสำหรับอัตราการเรียนรู้

$J^{T}J$ คือ Hessian Matrix

I คือ เมตริกซ์เอกลักษณ์ (Identity Matrix)

$J^{T}e$ คือ gradient โดยที่ e คือเวคเตอร์ความคลาดเคลื่อนของโครงข่าย

การหยุดการสอน (Stop training)

การหยุดการสอนโครงข่ายสามารถกระทำได้ 2 วิธี คือ 1) การกำหนดรอบการสอน (Epochs) 2) การกำหนดค่าความคลาดเคลื่อน (error) ที่ยอมรับได้ (Bhokha, 1998) Carpenter (1993) ได้แนะนำให้กำหนดจำนวนรอบการสอนอยู่ที่ 20,000 ถึง 100,000 รอบ และอีกวิธี คือ การกำหนดค่าความคลาดเคลื่อนระหว่างข้อมูลจริง และผลลัพธ์ที่โครงข่ายสามารถคำนวณได้ (Khan et al. 1993 : 81-94) แต่ข้อควรระวังก็คือ การสอนโครงข่ายที่นานเกินไปจะทำให้เกิดปัญหาที่เรียกว่า Over fitting ได้ ดังภาพที่ 6 ซึ่งคือปัญหาที่โครงข่ายสามารถที่จะเรียนรู้จนได้ผลลัพธ์ที่มีค่าความคลาดเคลื่อน ที่น้อยที่สุด แต่เมื่อนำมาตรวจสอบความถูกต้องด้วยชุดทดสอบแล้ว ปรากฏว่าไม่สามารถที่จะให้ผลลัพธ์ได้ดีจริง (Bhokha,1998)

ภาพที่ 6 กรณี over fitting

ที่มา : ศรัลย์ ปานศรีพงษ์, “การพัฒนาแบบจำลองโครงข่ายใยประสาทประดิษฐ์และแบบจำลองความถดถอยเชิงพหุ เพื่อทำนายความเข้มข้นของ PM10 ในพื้นที่กรุงเทพมหานครและเทศบาลนครราชสีมา” (วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชา วิศวกรรมสิ่งแวดล้อม บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีสุรนารี, 2550), 21.

ชุดข้อมูล (Samples)

ชุดข้อมูล คือข้อมูลที่ทราบตัวแปรต้นและตัวแปรตามเพื่อนำมาใช้สอนโครงข่าย Yeh et al. (1993) ได้กล่าวไว้ว่า แหล่งที่มาของชุดข้อมูลแบ่งออกได้เป็น 3 ลักษณะ คือ 1) แบบสอบถาม 2) ข้อมูลทางสถิติ และ 3) จากการทดลอง โดยชุดข้อมูลดังกล่าวจะนำมาแบ่งออกเป็น 2 ชุด คือ ชุดการสอน (Training set) และ ชุดทดสอบ (Test set) Klimasauskas (1993) ได้ให้ความเห็นว่า ควรให้มีจำนวนชุดการสอนอย่างน้อย 5 ชุด เพื่อใช้ในการสอนโครงข่าย

การทดสอบโครงข่าย (Testing)

Smith (1993) กล่าวว่า การทดสอบโครงข่ายเป็นการทดสอบว่าโครงข่ายสามารถที่จะเรียนรู้จากชุดการสอน (Training set) ได้ดีเพียงไร โดยใช้ชุดข้อมูลที่ไม่เคยใช้สำหรับการสอนมาทดสอบ เรียกว่า ชุดทดสอบ (Test set) ซึ่งโครงข่ายที่สามารถให้ผลลัพธ์ที่แม่นยำได้เมื่อใช้ชุดทดสอบมาทดสอบ จะเป็นโครงข่ายที่น่าเชื่อถือ ในการทดสอบโครงข่ายสามารถแบ่งออกได้เป็น 2 ลักษณะ คือ

- แบ่งข้อมูลออกเป็น 2 ชุด (McKim et al. 1996 : 247-256) โดยชุดแรกไว้สำหรับสอนโครงข่ายให้จดจำรูปแบบของข้อมูล และข้อมูลชุดที่สองไว้สำหรับทดสอบโครงข่าย โดยผลลัพธ์ที่แตกต่างระหว่างค่าจริง และค่าที่ได้จากชุดทดสอบจะถูกคำนวณออกมาเป็นค่าผิดพลาดของระบบ (system error) ซึ่งค่าผิดพลาดของระบบที่น้อย จะแสดงถึงความสามารถในการทำนายที่สูง
- ใช้ชุดข้อมูลทั้งหมดเป็นทั้งชุดการสอน และชุดทดสอบ โดยนำชุดข้อมูลทั้งหมดมาสอนโครงข่ายก่อน และหลังจากนั้นนำชุดข้อมูลชุดเดิมมาทดสอบโครงข่าย

วิธีจีเนติกอัลกอริทึม

จีเนติกอัลกอริทึม (Genetic algorithm: GA) (อาทิตย์ ศรีแก้ว 2002 : 69-83) เป็นวิธีการค้นหาคำตอบที่ดีที่สุดโดยใช้หลักการคัดเลือกแบบธรรมชาติและหลักการทางสายพันธุ์ จีนเนติกอัลกอริทึมเป็นการคำนวณอย่างหนึ่งที่กล่าวได้ว่ามี “วิวัฒนาการ” อยู่ในขั้นตอนของการหาคำตอบ และได้รับการจัดให้เป็นวิธีหนึ่งในกลุ่มของการคำนวณเชิงวิวัฒนาการ ซึ่งปัจจุบันเป็นที่ยอมรับในประสิทธิภาพ และมีการนำไปประยุกต์ใช้อย่างกว้างขวางในงานทางด้านปัญญาประดิษฐ์ต่างๆ

จีเนติกอัลกอริทึมเป็นขั้นตอนในการค้นหาคำตอบให้ระบบ เราสามารถมอง GA เป็นเครื่องมือในการช่วยคำนวณอย่างหนึ่ง โดยธรรมชาติแล้วประกอบไปด้วย 3 กระบวนการที่สำคัญ ได้แก่

1. การคัดเลือกสายพันธุ์ (Selection) คือขั้นตอนในการคัดเลือกประชากรที่ดีในระบบไปเป็นต้นกำเนิดในสายพันธุ์เพื่อให้กำเนิดลูกหลานในรุ่นต่อไป

2. ปฏิบัติการทางสายพันธุ์ (Genetic operation) คือกรรมวิธีการเปลี่ยนแปลงโครโมโซมด้วยวิธีการทางสายพันธุ์ เป็นขั้นตอนการสร้างลูกหลานซึ่งได้จากการรวมพันธุ์ของต้นกำเนิดสายพันธุ์เพื่อให้ลูกหลานที่มีส่วนผสมผสานมาจากพ่อแม่หรือได้จากการแปรผันยีนของพ่อแม่เพื่อให้ได้ลูกหลานสายพันธุ์ใหม่เกิดขึ้น

3. การแทนที่ (Replacement) คือขั้นตอนการนำอาลูกหลานกำเนิดใหม่ไปแทนที่ประชากรเก่าในรุ่นก่อน เป็นขบวนการในการคัดเลือกว่าควรจะนำลูกหลานในกลุ่มใดไปแทนประชากรเก่าในกลุ่มใด

ขั้นตอนการทำงานของจีเนติกอัลกอริทึม

คำตอบของระบบที่ต้องการให้ GA ทำการค้นหาจะอยู่ในรูปของโครโมโซมในกลุ่มของประชากรซึ่งคำตอบที่ต้องการจะเป็นโครโมโซมที่ดีที่สุดในกลุ่ม ดังนั้นระบบจะสามารถรู้ได้ว่าคำตอบที่มีอยู่ใน GA ณ เวลาหนึ่งๆ นั้นดีหรือไม่ดีอย่างไรด้วยการประเมินค่าของโครโมโซมผ่านฟังก์ชันวัตถุประสงค์ ขั้นตอนทั่วไปของ GA มีดังต่อไปนี้

1. สร้างประชากร โดยปกติจะใช้การสุ่ม (random)

2. ประเมินค่าโครโมโซมของกลุ่มประชากรทั้งด้วยฟังก์ชันวัตถุประสงค์ เนื่องจากระบบไม่สามารถเข้าใจค่าของโครโมโซมภายใน GA ดังนั้นโครโมโซมจะต้องผ่านการถอดรหัสก่อนที่จะนำไปทำการคำนวณด้วยฟังก์ชันวัตถุประสงค์ได้

3. คำนวณหาค่าความเหมาะสมแล้วส่งกลับไปยัง GA

4. ใช้ค่าความเหมาะสมทำการคัดเลือกโครโมโซมบางกลุ่มเพื่อนำมาเป็นต้นกำเนิดสายพันธุ์ซึ่งจะใช้เป็นตัวแทนในการถ่ายทอดสายพันธุ์ให้กับรุ่นถัดไป

5. นำต้นกำเนิดสายพันธุ์มาทำการสร้างลูกหลานด้วยปฏิบัติการทางสายพันธุ์ โครโมโซมที่ได้ในขั้นตอนนี้ก็คือโครโมโซมลูกหลาน

6. คำนวณค่าความเหมาะสมของโครโมโซมลูกหลานโดยใช้ขั้นตอนเดียวกับข้อ 3

7. โครโมโซมในประชากรเดิมจะถูกแทนที่ด้วยลูกหลานที่ได้จากข้อ 5 ประชากรเพียงบางส่วนเท่านั้นที่จะถูกแทนที่ด้วยกลวิธีเฉพาะสำหรับขั้นตอนของการแทนที่โดยใช้ค่าความเหมาะสมในการตัดสิน

8. เริ่มต้นทำซ้ำจากขั้นตอนในข้อ 2 ไปเรื่อยๆจนกระทั่งได้คำตอบที่ต้องการ คำตอบที่ได้จะมาจากโครโมโซมที่ดีที่สุดในกลุ่มประชากรนั่นเองโดยที่สามารถใช้ค่าจากฟังก์ชันวัตถุประสงค์เพื่อเป็นการประเมินว่าคำตอบที่ได้เป็นที่ต้องการแล้วหรือไม่

องค์ประกอบที่สำคัญของจีเนติกอัลกอริทึม

มีองค์ประกอบที่สำคัญทั้งหมด 5 ส่วนด้วยกันดังต่อไปนี้

1. การเข้ารหัสโครโมโซม (Chromosome encoding) คือขั้นตอนสำหรับแปลงทางเลือกสำหรับการแก้ปัญหาที่เป็นไปได้ให้อยู่ในรูปแบบของโครโมโซม (Chromosome)

2. ประชากรต้นแบบ (Initial Population) คือการสุ่มเลือกเพื่อสร้างประชากรต้นแบบขึ้นมาเพื่อใช้เป็นจุดเริ่มต้นของขั้นตอนการวิวัฒนาการ

3. ฟังก์ชั่นวัตถุประสงค์ (Fitness Function) คือฟังก์ชั่นสำหรับประเมินค่าความเหมาะสม เพื่อให้คะแนนสำหรับคำตอบต่างๆ ที่เป็นไปได้ของปัญหา

4. ปฏิบัติการทางสายพันธุ์ (Genetic operation) คือการดำเนินการต่างๆ ตามขั้นตอนของ GA เพื่อให้การเกิดวิวัฒนาการไปสู่คำตอบที่ดีขึ้น ซึ่งได้แก่ การคัดเลือกสายพันธุ์ (Selection), ครอสโอเวอร์ (Crossover) และมิวเทชั่น (Mutation)

5. พารามิเตอร์ (Parameter) คือปัจจัยที่ส่งผลต่อการทำงานของ GA เช่น ขนาดของประชากร, ความน่าจะเป็นของการครอสโอเวอร์ หรือความน่าจะเป็นของการมิวเทชั่น

การเข้ารหัสโครโมโซม (Chromosome Encoding)

การเข้ารหัสประชากรเป็นขั้นตอนแรกและเป็นขั้นตอนที่สำคัญเพราะเป็นการออกแบบให้โครโมโซมเป็นตัวแทนของคำตอบจากระบบ ชุดคำตอบของระบบ แต่ละระบบจะมีจำนวนตัวแปรไม่

เท่ากัน ขึ้นอยู่กับลักษณะของปัญหา ความซับซ้อนและการออกแบบแก้ปัญหาของระบบนั้นๆ ในการแปลงวิธีการสำหรับแก้ปัญหาที่เป็นไปได้ ให้อยู่ในรูปแบบของ โครโมโซม มีหลายวิธีได้แก่

Permutation Encoding เป็นรูปแบบ Permutation encoding สามารถใช้แก้ปัญหาที่เป็นแบบลำดับได้ เช่นปัญหา travelling salesman หรือปัญหาที่เกี่ยวกับลำดับการทำงาน ใน permutation encoding ทุกโครโมโซมเป็นจำนวนนับ (string number) ที่แทนจำนวนลำดับการทำงาน ยกตัวอย่างเช่น

โครโมโซม A : 1 5 3 2 6 4 7 9 8

โครโมโซม B : 8 5 6 7 2 3 1 4 9

รูปแบบนี้ถูกใช้มากในการแก้ปัญหาแบบมีลำดับ ซึ่งจำเป็นที่จะต้องสร้าง Crossover และ mutation ให้มีความสอดคล้องกับโครโมโซม ปัญหาที่ใช้รูปแบบโครโมโซมแบบนี้ในการแก้ปัญหาเช่น ปัญหาของ Travelling Salesman

Valued Encoding หรือเรียกว่า Direct Encoding ทุกตำแหน่งของโครโมโซมจะมีค่าบางค่าซึ่งสามารถเชื่อมโยงไปยังปัญหาได้ เช่นตัวอักษร จำนวนจริง คำสั่งหรืออื่นๆ รูปแบบของ valued encoding สามารถใช้กับปัญหาที่มีความซับซ้อนของค่าที่ รูปแบบของ binary encoding ไม่สามารถใช้ได้ ตัวอย่างเช่น

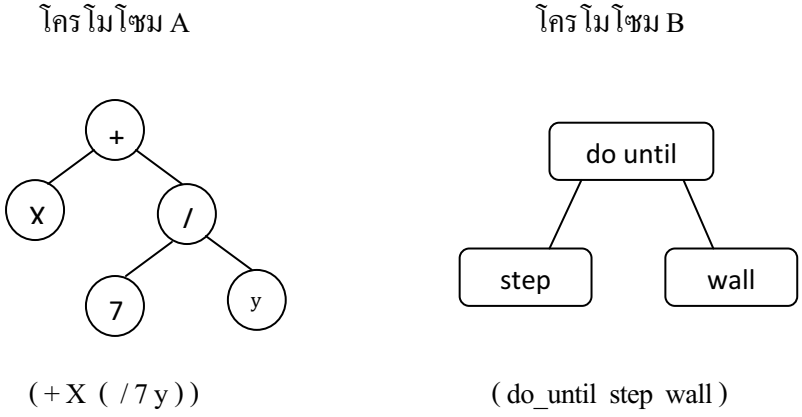
โครโมโซม A : 1.2324 5.3243 0.4556 2.3293 2.4545

โครโมโซม B : ABCSGTHIMNASBVFRTOIUCEUDSAG

โครโมโซม C : (back), (back), (right), (left), (left)

รูปแบบ valued encoding เป็นวิธีการที่ดีสำหรับปัญหาพิเศษบางอย่าง แต่รูปแบบนี้จำเป็นที่ต้องพัฒนาการครอสโอเวอร์และมิวเทชั่น ในแต่ละปัญหาด้วย ปัญหาที่ใช้รูปแบบโครโมโซมแบบนี้ในการแก้ปัญหาเช่น ปัญหาของการหาค่าน้ำหนักในโครงข่ายประสาทเทียม (neural network)

Tree Encoding รูปแบบนี้ใช้แก้ปัญหาสำหรับการพัฒนาโปรแกรมสำหรับ GA ทุกโครโมโซมจะเป็น โหนด (node) ของต้นไม้ เช่นฟังก์ชั่นหรือคำสั่งในภาษา LISP ยกตัวอย่างดังภาพที่ 7



ภาพที่ 7 การเข้ารหัสโครโมโซมแบบต้นไม้

ประชากรต้นแบบ (Initial Population)

ขั้นตอนนี้จะเป็นขึ้นตอนแรกที่เกิดขึ้นก่อนที่จะเริ่มเข้ากระบวนการของ GA โดยประชากรกลุ่มแรก หรือประชาการต้นกำเนิด จะเกิดจากการสุ่มเลือกขึ้นมาจาก กลุ่มของประชากรทั้งหมดที่มีอยู่ โดยในการสุ่มเลือกจะทำการสุ่มตามจำนวนของประชากรที่ได้กำหนดไว้เป็นพารามิเตอร์ของอัลกอริทึม

ฟังก์ชั่นวัตถุประสงค์ (Fitness Function)

โครโมโซมทุกตัวจะมีค่าความเหมาะสมของตัวเองเพื่อใช้สำหรับพิจารณาว่าโครโมโซมตัวนั้น เหมาะหรือไม่ที่จะนำมาใช้สืบทอดพันธุกรรมสำหรับสร้างโครโมโซมรุ่มใหม่ โดยวิธีการสำหรับคิดค่าความเหมาะสมนั้นจะใช้สมการที่สอดคล้องกับแต่ละปัญหา

ปฏิบัติการทางสายพันธุ์ (Genetic operation)

ปฏิบัติการทางสายพันธุ์ (Genetic operation) นั้นถือได้ว่าเป็นหัวใจหลักที่สำคัญของ GA โดยการกระบวนการพื้นฐานที่สำคัญ 3 ขั้นตอนคือ

1. การคัดเลือกสายพันธุ์ (Selection)

ในการคัดเลือกโครโมโซมขึ้นมาจากกลุ่มประชากรเพื่อใช้สำหรับสืบทอดพันธุกรรมให้กับลูกหลานนั้นมีความสำคัญมากเนื่องจากในการคัดเลือกควรที่จะเลือกโครโมโซมที่

มีความเหมาะสมที่จะสืบทอดพันธุกรรมเท่านั้น จึงมีการพัฒนารูปแบบต่างๆ ที่ใช้สำหรับคัดเลือกโครโมโซม ได้แก่

การคัดเลือกแบบวงล้อรูเล็ต (Roulette Wheel) เป็นการคัดเลือกแบบการให้น้ำหนักค่าความเหมาะสมกับแต่ละโครโมโซม โดยโครโมโซมที่มีค่าความเหมาะสมมากย่อมมีโอกาสถูกเลือกได้มาก และโครโมโซมที่มีค่าความเหมาะสมน้อยๆ ก็จะมีโอกาสถูกเลือกได้น้อย หรืออาจไม่ได้รับการคัดเลือกเลย

การคัดเลือกแบบให้ลำดับ (ranking) คือจะทำการเลือกประชากรที่มีค่าความเหมาะสมที่ดีที่สุด โดยไม่สนใจประชากรตัวอื่นเลย

การคัดเลือกแบบอีลิททิซึม (Elitism) โดยวิธีนี้จะทำการคัดลอกประชากรที่มีค่าความเหมาะสมที่ดีที่สุดเข้าไปในประชากรรุ่นต่อไปก่อน จากนั้นค่อยใช้วิธีการเลือกในแบบอื่นๆ เพื่อทำการคัดเลือกโครโมโซมตัวต่อๆ ไป

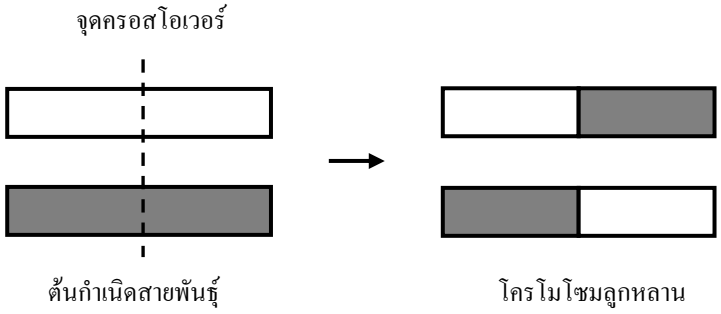
การคัดเลือกแบบวิธีการแข่งขัน (Tournament Selection) เป็นวิธีการเดียวกับการแข่งขันกีฬาทั่วๆไป ทำได้โดยการสุ่มแบ่งกลุ่มคัดเลือกโครโมโซมแล้วเลือกเอาโครโมโซมที่ดีที่สุดในกลุ่มนั้นเพื่อเป็นต้นกำเนิดสายพันธุ์ จำนวนของโครโมโซมในแต่ละกลุ่มนั้นจะแตกต่างกันออกไป โดยปกติแล้วจะใช้วิธีสุ่มแบบจับคู่โครโมโซม (นั่นคือมีเพียง 2 โครโมโซมที่ถูกสุ่มเลือกเข้ามาในแต่ละการแข่งขัน) วิธีการจัดการแข่งขันมีความเหมาะสมในการทำให้ปัญหาความเหลื่อมล้ำของค่าความเหมาะสมของโครโมโซมหมดไป

2. ครอสโอเวอร์ (Crossover)

ครอสโอเวอร์ (Crossover) เป็นวิธีการรวมตัวใหม่ของโครโมโซม (Recombination operator) โดยทำการรวมส่วนย่อยระหว่างโครโมโซมต้นกำเนิดสายพันธุ์ตั้งแต่สองโครโมโซมขึ้นไป เพื่อให้กลายเป็นโครโมโซมลูกหลาน โครโมโซมลูกหลานที่ได้จากการครอสโอเวอร์นี้จะมีพันธุกรรมจากต้นกำเนิดสายพันธุ์อยู่ในตัว โดยปกติทั่วไปแล้วจะมีการกำหนดอัตราการทำครอสโอเวอร์เอาไว้ ซึ่งส่วนใหญ่จะใช้ความน่าจะเป็น (P_c) เป็นตัวกำหนดอัตราดังกล่าว วิธีการทำครอสโอเวอร์มีได้หลายแบบดังรายละเอียดในตัวอย่างต่อไปนี้

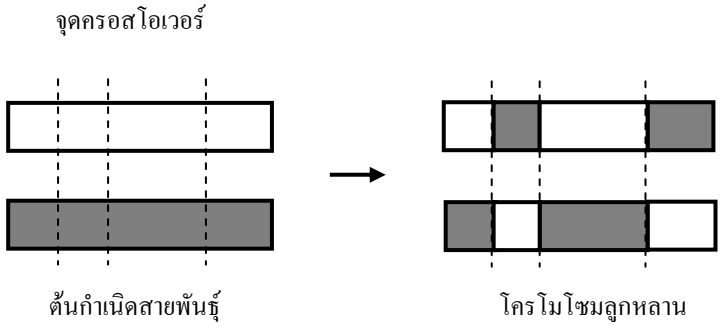
การทำครอสโอเวอร์แบบจุดเดียว (Single-Point Crossover) การทำครอสโอเวอร์แบบจุดเดียวนี้ โครโมโซมลูกหลานจะมีสายพันธุ์ของต้นกำเนิดอยู่อย่างละหนึ่งส่วน จุดตัดในการทำครอส

โอเวอร์นั้นโดยปกติจะได้มาจากการสุ่มเลือก ตัวอย่างของการทำ ครอสโอเวอร์แบบจุดเดียว แสดงในภาพที่ 8



ภาพที่ 8 ตัวอย่างของ one-point crossover
ที่มา : ปรับปรุงจาก อาทิยต์ ศรีแก้ว, “จีเนติกอัลกอริทึม ตอนที่ 1,” วารสารเทคโนโลยีสุรนารี 9, 1 (มกราคม-มีนาคม 2545) : 69-83.

การทำครอสโอเวอร์แบบสองจุด (Two-Point Crossover) ตัวอย่างการทำครอสโอเวอร์แบบหลายจุดดังแสดงในรูปที่ 10 มีการใช้จุดตัดทั้งหมด 3 จุด ดังนั้นโครโมโซมลูกหลานจะมีสายพันธุ์ของต้นกำเนิดอยู่มากกว่าหนึ่งส่วน หลักการเลือกจุดของครอสโอเวอร์นั้นมีอยู่หลายแบบ แต่ละแบบจะให้ผลต่อการเปลี่ยนแปลงของสายพันธุ์ในโครโมโซมลูกหลานที่แตกต่างกันออกไปด้วย วิธีที่ง่ายและเป็นที่นิยมใช้ทั่วไปคือการสุ่มเลือกจุดครอสโอเวอร์ การทำครอสโอเวอร์หลายจุดจะให้ผลของลูกหลานที่มีความหลากหลายกว่าการทำ ครอสโอเวอร์แบบจุดเดียว อันจะมีผลให้การลู่เข้าสู่คำตอบของระบบสามารถครอบคลุมพื้นที่ของคำตอบได้มากยิ่งขึ้น อย่างไรก็ตามการทำครอสโอเวอร์แบบหลายจุด ซึ่งทำให้เกิดการเปลี่ยนแปลงของโครโมโซมลูกหลานได้มากกว่าการทำครอสโอเวอร์แบบจุดเดียวนั้นอาจทำให้โอกาสเบี่ยงเบนของคำตอบที่มีอยู่ในโครโมโซมลูกหลานมีอัตราที่สูงกว่าเช่นกัน การทำครอสโอเวอร์ที่นิยมอีกชนิดหนึ่งคือ ครอสโอเวอร์แบบสม่ำเสมอ (Uniform Crossover) ซึ่งมีข้อดีในการลดปัญหาความไม่สมดุลในการแบ่งจุดครอสโอเวอร์แบบจุดเดียวของโครโมโซมที่ขนาดต่างๆกัน แสดงในภาพที่ 9

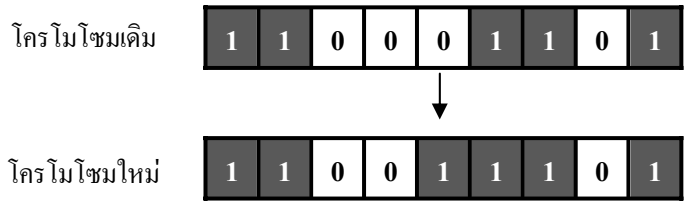


ภาพที่ 9 ตัวอย่างของ two-point crossover

ที่มา : ปรับปรุงจาก อาทิตย์ ศรีแก้ว, “จีเนติกอัลกอริทึม ตอนที่ 1,” วารสารเทคโนโลยีสุรนารี 9, 1 (มกราคม-มีนาคม 2545) : 69-83.

3. มิวเทชั่น (Mutation)

มิวเทชั่น (Mutation) เป็นวิธีการแปรผันยีนหรือส่วนย่อยของโครโมโซม ซึ่งสามารถเปรียบเทียบได้กับการกลายพันธุ์ของสิ่งมีชีวิตในทางชีววิทยานั่นเอง มิวเทชั่นคือ การเปลี่ยนแปลงยีนในโครโมโซมซึ่งในทางปฏิบัติแล้วยีนก็คือบิตในระบบตัวเลขของคอมพิวเตอร์ แสดงในภาพที่ 10



ภาพที่ 10 ตัวอย่างของการมิวเทชั่น

ที่มา : ปรับปรุงจาก อาทิตย์ ศรีแก้ว, “จีเนติกอัลกอริทึม ตอนที่ 1,” วารสารเทคโนโลยีสุรนารี 9, 1 (มกราคม-มีนาคม 2545) : 69-83.

การทำมิวเทชั่นเปรียบเสมือนกับการก้าวเดินไปสู่คำตอบของระบบเช่นเดียวกับการทำครอสโอเวอร์ นอกเหนือจากนั้นแล้วมิวเทชั่นยังสามารถมองเป็นการทำให้เกิดความหลากหลายขึ้นในกลุ่มประชากร มีผลให้คำตอบที่เกิดขึ้นในขบวนการของ GA ครอบคลุมพื้นที่การค้นหาคำตอบทั่วถึงยิ่งขึ้น อย่างไรก็ตามอัตราในการทำมิวเทชั่นเป็นปัจจัยที่สำคัญอีกอย่างหนึ่งที่ต้องคำนึงถึงเพราะจะมี

ผลต่อพฤติกรรมการทำงานของ GA มีผลการค้นคว้ารายงานว่า อัตราการทำมิวเทชั่นจะขึ้นอยู่กับขนาดของประชากร เพื่อให้การสำรวจพื้นที่ในการหาคำตอบเป็นไปอย่างทั่วถึง ดังนั้นการกำหนดอัตราการทำมิวเทชั่นต้องมีความเหมาะสมที่สุดต่อระบบด้วยเพื่อก่อให้เกิดผลในการค้นหาคำตอบอย่างมีประสิทธิภาพยิ่งขึ้น ซึ่งโดยปกติแล้วการทำมิวเทชั่นจะมีอัตราการใช้งานที่ค่อนข้างต่ำ

พารามิเตอร์ (Parameter)

พารามิเตอร์ (Parameter) ที่สำคัญหรือเป็นพื้นฐานของ GA มี 3 ตัวได้แก่

ความน่าจะเป็นของการครอสโอเวอร์ (Crossover Probability) มีค่าอยู่ในช่วง 0-100 จากการทดลอง จากการทดลองของนักวิทยาศาสตร์หลายท่าน พบว่าความน่าจะเป็นของการ crossover ส่วนใหญ่อยู่ที่ 60-95% ถ้าพบว่าไม่มีการครอสโอเวอร์ (0%) ผลที่ได้คือการทำสำเนาโครโมโซมจากรุ่นพ่อแม่ (parent) แต่ถ้ามีการครอสโอเวอร์บ่อยๆ จะทำให้เกิดผลลัพธ์ที่หลากหลาย บางปัญหาอาจจะมีอัตราการครอสโอเวอร์ที่ 60% ก็จะให้ผลลัพธ์ดีที่สุด

ความน่าจะเป็นของการมิวเทชั่น (Mutation Probability) มีค่าอยู่ในช่วง 0-100 จากการทดลองของนักวิทยาศาสตร์หลายท่าน พบว่าความน่าจะเป็นของการมิวเทชั่น ส่วนใหญ่อยู่ที่ 0.5-1% ซึ่งอัตราการเกิดมิวเทชั่น ไม่เกิดบ่อยมากนัก เนื่องจากจะทำให้วิธีการแบบ GA เปลี่ยนไปเป็นวิธี random search

ขนาดของประชากร (Population size) คือขนาดของประชากรถ้ามีจำนวนมากจะทำให้การประมวลผลในการค้นหาคำตอบในกระบวนการ GA ช้าลง ขนาดของประชากรที่เหมาะสมประมาณ 20-30 แต่บางครั้งจำนวนประชากรที่เหมาะสมประมาณ 50-100 ซึ่งจากบางงานวิจัยแสดงให้เห็นว่าขนาดประชากรที่เหมาะสมจะขึ้นอยู่กับการถอดรหัสของโครโมโซมด้วย

เงื่อนไขในการหยุดกระบวนการค้นหาคำตอบ

ในการหยุดกระบวนการค้นหาคำตอบของกระบวนการ GA นั้นมีได้หลายวิธีเช่น

- ครบจำนวนรอบที่ได้กำหนดไว้
- พบเป้าหมายหรือคำตอบที่ต้องการแล้ว
- พบคำตอบที่ใกล้เคียงกับคำตอบที่ต้องการ เช่น โครโมโซมมีค่า ณ ตำแหน่งของยีนเดียวกันหรือเหมือนกันถึงร้อยละ 95

งานวิจัยที่เกี่ยวข้อง

สิทธิชัย เทพไพฑูรย์ (2545) ทำการศึกษาการทำนายสัญญาณการซื้อขายหลักทรัพย์ด้วยการใช้จีนเนติกอัลกอริทึมร่วมกับโครงข่ายประสาทเทียม ซึ่งถ้าได้ตัวทำนายที่ดีจะทำให้ได้ผลตอบแทนการลงทุนมหาศาลในเวลาอันรวดเร็ว โดยเสนอแบบจำลองใหม่ที่ไม่ใช้เทคนิคการวิเคราะห์การถดถอยและการวิเคราะห์อนุกรมเวลา เนื่องจากเครื่องมือเหล่านี้เป็นการหาความสัมพันธ์เชิงเส้นโดยที่ข้อมูลราคาหลักทรัพย์ไม่ใช่ข้อมูลเชิงเส้น และการทำนายโดยอาศัยโครงข่ายประสาทเทียมและการหาพารามิเตอร์ที่เหมาะสมมาเป็นข้อมูลที่ใช้ฝึกโครงข่ายประสาทเทียมนั้นไม่สามารถให้ผลลัพธ์ที่น่าพอใจได้เนื่องจาก ความแม่นยำในการพยากรณ์นั้นขึ้นกับข้อมูลที่จะนำมาฝึกสอนมากกว่าสถาปัตยกรรมของโครงข่ายประสาทเทียม การแบ่งข้อมูลว่าส่วนใดจะนำไปเป็น ชุดฝึกสอน (Training set) ชุดประเมินผล (Validation set) และชุดทดสอบ (Test set) ที่แตกต่างกัน ทำให้ความสามารถในการพยากรณ์ต่างกัน งานวิจัยนี้จึงได้เสนอแนวคิดใหม่โดยการนำวิธีจีนเนติกอัลกอริทึม มาใช้เลือกชุดฝึกสอนที่เหมาะสมให้กับโครงข่ายประสาทเทียม โดยใช้โครงข่ายประสาทเทียมแบบกระจายความผิดพลาดย้อนหลัง(Back prorogation Neural Network) 3 ชั้น แต่ละชั้นมีโหนด 6 3 1 ตามลำดับใช้ฟังชั่นแทนซิก (tansig) ทำหน้าที่เป็นทรานสเฟอร์ฟังก์ชั่น (transfer function)

ผลการศึกษาพบว่าการทำนายเมื่อใช้จีนเนติกอัลกอริทึมมาช่วยเลือกข้อมูล จะให้ค่าความถูกต้องน้อยกว่าการใช้วิธีเทรนแบบปกติแต่สามารถทำกำไรได้มากกว่า และผลการทำนายเมื่อใช้จีนเนติกอัลกอริทึมนั้นมีความถูกต้องแม่นยำในแต่ละช่วงเวลากระจายตัวได้ดีกว่า การใช้จีนเนติกอัลกอริทึม เลือกชุดข้อมูลที่จะนำมาเทรนนั้น ช่วยให้การทำนายมีความแม่นยำสม่ำเสมอมากขึ้นในแต่ละช่วงเวลา แม้ว่าค่าเฉลี่ยโดยรวมจะไม่สูงมากนัก แต่การที่ความสามารถของตัวทำนายคงที่จะช่วยให้ไม่ทำนายผิดต่อเนื่องเป็นเวลายาวนานมากเกินไป ช่วยลดโอกาสการขาดทุนเมื่อทำนายผิดได้

Somsong et al. (1997) ทำการศึกษาเรื่องวิธีจีเนติกอัลกอริทึมในการพยากรณ์ธุรกิจเงินฝาก ซึ่งงานวิจัยนี้นำเสนอการประยุกต์ใช้จีเนติกอัลกอริทึมในการพยากรณ์ธุรกิจเงินฝากโดยการหารูปแบบของความสัมพันธ์ระหว่างปัจจัยนำเข้าคือ 1) ผลผลิตมวลรวมของประชาชาติ (GDP) 2) อุปสงค์การเงิน (money supply) 3) ดอกเบี้ย (interest rate) 4) จำนวนสาขาของธุรกิจเงินฝาก (number of branches of commercial banks) และ 5) เงินกู้ (loan) และผลลัพธ์คือเงินฝากธนาคาร (bank deposit) โดยการคำนวณในกระบวนการของ GA ประกอบไปด้วย 2 loop

The genetic forecasting loop เป็นวิธีในการ minimize error ระหว่างค่าจริงกับค่าพยากรณ์ และ minimize average pattern error

Pattern learning loop เป็นวิธีในการเรียนรู้รูปแบบความสัมพันธ์ที่ทำให้กำลังในการพยากรณ์ของอัลกอริทึมให้สูงขึ้น

โดยพารามิเตอร์ของ GA ของงานวิจัยนี้คือ

Population size: 30

Crossover rate: 1

Mutation rate: 0.01

Number of crossing sites: 1

Number of best individuals passed on the next generation: 2

เปรียบเทียบกับวิธีโครงข่ายประสาทเทียม โดยมี 3 ชั้น ชั้นนำเข้า (input node) 5 โหนด ชั้นซ่อน 7 โหนด และชั้นผลลัพธ์ 1 โหนด ตามลำดับ

ผลลัพธ์จากการศึกษาพบว่า เมื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์ด้วยเปอร์เซ็นต์ความคลาดเคลื่อนเฉลี่ย พบว่าโมเดล ANN ให้ค่าเปอร์เซ็นต์ความคลาดเคลื่อนเฉลี่ยเท่ากับ 8.29 และโมเดล GA ให้เปอร์เซ็นต์ความคลาดเคลื่อนเฉลี่ยเท่ากับ 5.07 ซึ่งน้อยกว่าวิธี ANN ดังนั้นสรุปได้ว่าโมเดลจากวิธี GA ให้ผลลัพธ์ในการพยากรณ์ดีกว่า

Ingleby and Crowe (2001) ทำการศึกษาเรื่อง โมเดลโครงข่ายประสาทเทียมสำหรับพยากรณ์อินทรีย์วัตถุของดิน จากเมือง Saskatchewan ประเทศแคนาดา โดยเป้าหมายของการศึกษาเพื่อเปรียบเทียบประสิทธิภาพของ โมเดลการถดถอยเชิงพหุคูณกับโมเดลโครงข่ายประสาทเทียม โดยมีวัตถุประสงค์คือ 1) เรียนรู้โครงข่ายในการพยากรณ์อินทรีย์คาร์บอน โดยใช้ข้อมูลที่กำหนดความยาวคลื่นที่แตกต่างกัน 2) คำนวณประสิทธิภาพโดยใช้ค่า SSE โครงข่ายประสาทเทียมใช้สถาปัตยกรรมแบบ multi-layer feedforward โดยประกอบด้วยอินพุต 5 โหนด จำนวนชั้นซ่อน 10 โหนด และในชั้นเอาท์พุต 1 โหนด โดยระหว่างชั้นอินพุตกับชั้นเอาท์พุต ใช้ log-sigmoid transfer function ระหว่างชั้นซ่อนกับชั้นเอาท์พุตใช้ linear transfer function การเลือกจำนวนโหนดในชั้นซ่อนตั้งแต่ 5, 10, 15 และ 20 โหนดตามลำดับ และใช้อัลกอริทึมในการเรียนรู้แบบ Levenberg-Maquadt back propagation เนื่องจากเป็นวิธีที่ให้ผลลัพธ์รวดเร็วในการเรียนรู้ข้อมูลที่มีขนาดปานกลาง และมีประสิทธิภาพ

ผลลัพธ์ที่ได้เปรียบเทียบกับผลลัพธ์ของโมดลที่ได้จากการทดสอบในงานวิจัยก่อนหน้าของ Ingleby และ Crowe (2000) เรื่องโมเดลการสะท้อนของเสปคตรัม สำหรับการพยากรณ์อินทรีย์คาร์บอน ของดินในเมือง Saskatchewan เช่นเดียวกัน

ผลการศึกษาพบว่า โมเดลโครงข่ายประสาทเทียมที่ให้ค่า SSE ดีที่สุดคือ รูปแบบ Outlook ที่มี 4 โหนดในชั้นอินพุต มีค่า SSE เท่ากับ 0.67 ในชุดประเมินผล โมเดลการถดถอยเชิงพหุคูณ ที่ให้ค่า SSE ดีที่สุดคือ รูปแบบ Outlook ที่มี 4 โหนดในชั้นอินพุต มีค่า SSE เท่ากับ 1.56 ซึ่งผลสรุปโดยภาพรวมแล้ว โมเดลที่มีประสิทธิภาพมากกว่าคือโมเดลโครงข่ายประสาทเทียม

Merdun et al. (2005) ทำการศึกษาการเปรียบเทียบวิธีโครงข่ายประสาทเทียมและการถดถอย สำหรับฟังก์ชันการแปลงค่าสำหรับการเก็บรักษาน้ำในดินและความสามารถในการทำให้น้ำซึมผ่าน พารามิเตอร์ที่ใช้ในการประมาณค่าของคุณสมบัติของดินมาจากคุณสมบัติพื้นฐานของดิน เช่น การกระจายของขนาดของดิน (particle-size distribution) ความหนาแน่นรวมของดิน (bulk density) และความพรุน (pore size) วิเคราะห์และประเมินผลโดยใช้วิธีโครงข่ายประสาทเทียม (artificial neural network) และการถดถอยเชิงเส้นพหุคูณ (multiple linear regression) เปรียบเทียบความสามารถของทั้งสองโมเดลโดยใช้เกณฑ์วัดประสิทธิภาพ โดยใช้ตัวอย่างดินในการวิเคราะห์ 195 ตัวอย่าง แบ่งเป็นสองกลุ่มคือตัวอย่างดิน 130 ตัวอย่างจะนำมาวิเคราะห์และพัฒนาผลลัพธ์ ตัวอย่างดิน 95 ตัวอย่างจะนำมาประเมินผลฟังก์ชันการแปลงค่า (pedotransfer functions : PTFs) ที่ได้ และเปรียบเทียบความแม่นยำของการทำนาย โดยใช้ค่าสัมประสิทธิ์การกำหนด (R^2) และค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE)

ผลการศึกษาพบว่า แม้ว่าสองวิธีที่ใช้จะไม่แตกต่างกันอย่างมีนัยสำคัญ ($p > 0.05$) แต่ผลลัพธ์ที่ได้จากการใช้วิธีการถดถอย ให้ผลลัพธ์ดีกว่าวิธีโครงข่ายประสาทเทียม และทั้งสองวิธีให้ความแม่นยำน้อย สำหรับวิธีการถดถอย ค่าของ R^2 อยู่ระหว่าง 0.637 ถึง 0.979 และค่าของ RMSE อยู่ระหว่าง 0.013 ถึง 0.938 และวิธีโครงข่ายประสาทเทียม ค่าของ R^2 อยู่ระหว่าง 0.444 ถึง 0.952 และค่าของ RMSE อยู่ระหว่าง 0.020 ถึง 3.551 ตามลำดับ แม้ว่าวิธีการถดถอยให้ผลลัพธ์ที่ดีกว่าในการศึกษานี้ แต่วิธีโครงข่ายประสาทเทียมก็เป็นประโยชน์ในการพัฒนาหรือการใช้เป็น algorithm ใหม่ในการพัฒนางานในอนาคต

Sousa et al. (2006) ทำการศึกษาการวิเคราะห์การถดถอยเชิงเส้นพหุคูณและโครงข่ายประสาทเทียมที่มีการเชื่อมโยงไปข้างหน้า ในการทำนายความเข้มข้นของโอโซน เพื่อใช้เตือนภัยที่เกิดจากระดับความเข้มข้นของโอโซนที่จะส่งผลเสียต่อมนุษย์ โดยใช้ตัวแปรความเข้มข้นของโอโซน (O_3), ไนโตรเจนมอนอกไซด์ (NO), ไนโตรเจนไดออกไซด์ (NO_2), ค่าเฉลี่ยของอุณหภูมิ (T), ความเร็วลม (WV) และความชื้นสัมพัทธ์ (RH) เป็นตัวแปรอิสระ และความเข้มข้นของโอโซนรายชั่วโมงในวันถัดไป เป็นตัวแปรตาม โดยเปรียบเทียบผลลัพธ์ที่ได้ก่อนที่จะใช้วิธีวิเคราะห์องค์ประกอบหลัก นั่นคือโมเดลการถดถอยเชิงเส้นพหุคูณ (Mulitple Linear Regression : MLR) และโมเดลโครงข่ายประสาทเทียมที่มีการเชื่อมโยงไปข้างหน้า (Feed-forward Artificial Neural Network : FANN) และเปรียบเทียบผลลัพธ์หลังจากการใช้วิธีวิเคราะห์องค์ประกอบหลักเพื่อลดปัจจัยนำเข้า นั่นคือโมเดลการถดถอยเชิงเส้นพหุคูณ (Principal Component Regression : PCR) และโมเดลโครงข่ายประสาทเทียมที่มีการเชื่อมโยงไปข้างหน้า (Principal Component Feed-forward Artificial Neural Network : PC-FANN) โดยพิจารณาตั้งแต่ 2-6 องค์ประกอบหลัก แล้วใช้ตัวชี้วัดประสิทธิภาพในการเปรียบเทียบโมเดล

ผลการศึกษาพบว่า ในการเปรียบเทียบวิธี MLR และ FANN วิธี FANN ให้ผลลัพธ์ดีกว่า และเมื่อใช้วิธีวิเคราะห์องค์ประกอบหลักเพื่อลดปัจจัยนำเข้าแล้ววิธี PC-FANN ให้ผลลัพธ์ดีกว่าวิธี PCR และให้ผลลัพธ์ดีที่สุดเมื่อพิจารณาที่ 5 องค์ประกอบหลัก และจากผลลัพธ์ทั้งหมดจะเห็นว่าการใช้วิธีวิเคราะห์องค์ประกอบหลักทำให้ผลลัพธ์ดีขึ้นเนื่องจากขจัดปัญหาความสัมพันธ์กันของตัวแปรอิสระ และการใช้วิธีโครงข่ายประสาทเทียมที่มีการเชื่อมโยงไปข้างหน้าให้ผลลัพธ์ดีกว่าวิธีการถดถอยเชิงเส้นพหุคูณเนื่องจากแก้ปัญหาความไม่เป็นเชิงเส้นของตัวแปร

Karunakar and Datta (2007) ศึกษาการควบคุมคุณสมบัติแบบทรายชื้นโดยใช้วิธีโครงข่ายประสาทเทียมและวิธีจีเนติกอัลกอริทึม ศึกษาเชิงเปรียบเทียบ วัตถุประสงค์ของการศึกษาเพื่อหาพารามิเตอร์การควบคุมที่เหมาะสมสำหรับแบบทรายชื้น โดยพิจารณา อินพุตจากคุณสมบัติแม่พิมพ์ คือ 1) แรงอัดในสภาวะชื้น (Green compress strength) 2) การลดแรงเฉือนในสภาวะเปียก (Green shear strength) 3) ความสามารถในการซึมผ่าน (Permeability) 4) แรงอัดสภาวะแห้ง (Dry compression strength) 5) การลดแรงเฉือนในสภาวะแห้ง (Dry shear strength) พารามิเตอร์การควบคุมคือเอาท์พุตที่ต้องการประกอบด้วย 1) จำนวน grain ที่เหมาะสม 2) เปอร์เซ็นต์ดินเหนียว 3) เปอร์เซ็นต์ความชื้น 4) Mulled time 5) Hardness โดยในการหาคำตอบในวิธีแรกคือวิธีโครงข่ายประสาทเทียมใช้ back

propagation algorithm ในการเรียนรู้ชุดข้อมูล ทำซ้ำจนกระทั่งคำตอบลู่เข้าสู่ค่าเอาท์พุตที่ต้องการ แล้วนำมาเปรียบเทียบกับชุดข้อมูลที่ได้จากการทดลองเมื่อกำหนดพารามิเตอร์อินพุตแล้ว วิธีที่สองคือวิธีจีเนติกอัลกอริทึม กำหนดขนาดประชากรเท่ากับ 5 ความน่าจะเป็นของการ crossover เท่ากับ 1.0 ความน่าจะเป็นของการ mutation เท่ากับ 0.0 และกำหนดเจนเนอชันเท่ากับ 81 โดย objective function วัดค่าชี้วัดประสิทธิภาพ (Performance Index : PI) โดยพัฒนาจากการใช้โมเดลการวิเคราะห์การถดถอย

ผลการศึกษาพบว่า จากการเปรียบเทียบผลลัพธ์ของโมเดลโครงข่ายประสาทเทียมกับโมเดลจีนเนติกอัลกอริทึม ด้วยการหาเปอร์เซ็นต์ความแตกต่างระหว่างผลการทดลองจริง จะเห็นว่าการใช้วิธีโครงข่ายประสาทเทียมให้ผลลัพธ์เป็นที่น่าพอใจ แต่มีความยุ่งยากในการทดสอบเพื่อเลือกสถาปัตยกรรมของโครงข่าย และการเรียนรู้ชุดข้อมูล ดังนั้นโดยภาพรวมแล้วผลลัพธ์จากวิธีจีนเนติกอัลกอริทึมให้ผลลัพธ์แม่นยำกว่า และสามารถให้ชุดพารามิเตอร์ที่เหมาะสมมากกว่า

Thomas et al. (2008) ได้ศึกษาการใช้วิธีจีเนติกอัลกอริทึมและวิธีการวิเคราะห์การถดถอยเชิงเส้นสำหรับการพยากรณ์ความต้องการของบ้านเช่าส่วนตัวในประเทศฮ่องกง โดยใช้ 4 โมเดลในการเปรียบเทียบประสิทธิภาพ แต่ละโมเดลจะสร้างโมเดลในการพยากรณ์สำหรับช่วงเวลา 5 ปี และ 10 ปี โดยทั้ง 4 โมเดลประกอบด้วย 1) โมเดลการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis: LRA) 2) โมเดลจีเนติกอัลกอริทึม (Genetic Algorithm: GA) 3) โมเดลจีเนติกร่วมกับการวิเคราะห์การถดถอยเชิงเส้น (GA-LRA) 4) โมเดลจีเนติกร่วมกับการวิเคราะห์การถดถอยเชิงเส้นแบบมีการปรับอัตราการมิวเทชั่น (GA-LRA with Adaptive Mutation Rate: GA-LRA with AMR) โดยตัวแปรอินพุตที่พิจารณาประกอบไปด้วยตัวชี้วัดทางเศรษฐกิจ 10 ตัวแปร ได้แก่ PUBH, LAND, UER, PROIN, HIS, GDP, GCON, HCPI, HSTOCK, GCE และค่าของเอาท์พุตที่ต้องการพยากรณ์คือ ความต้องการขายบ้านเช่าส่วนตัว โดยในวิธี LRA ใช้วิธี stepwise regression analysis วิธี GA-LRA จะใช้วิธีจีเนติกอัลกอริทึมในการเลือกพารามิเตอร์สำหรับการวิเคราะห์การถดถอยเชิงเส้น โดยขอบเขตของแต่ละพารามิเตอร์ได้จากค่าต่ำสุดและค่าสูงสุดจากพารามิเตอร์จากวิธี all possible regression

ผลลัพธ์จากการศึกษาพบว่า โมเดล LRA ทั้งการพยากรณ์ 5 ปีและ 10 ปีให้ผลลัพธ์ไม่เป็นที่น่าพอใจเมื่อเปรียบเทียบกับผลลัพธ์ที่ได้จากวิธี GA พบว่าให้ผลลัพธ์ในการพยากรณ์ดีกว่าโมเดล LRA ทั้งสองโมเดล ส่วนวิธี GA-LRA ก็ให้ผลลัพธ์ในการพยากรณ์ดีขึ้นจากโมเดล LRA

และ GA และผลลัพธ์จากวิธี GA-LRA with AMR เป็นวิธีสำหรับแก้ปัญหา local optima ซึ่งพบว่าเป็นโมเดลที่ให้ผลลัพธ์ดีที่สุด ซึ่งจากการเปรียบเทียบการพยากรณ์ทั้ง 4 โมเดลพบว่า 1) โมเดล LRA เป็นโมเดลที่ง่ายและใช้เวลาน้อยในการได้ผลลัพธ์ 2) วิธี GA เป็นวิธีที่ยากและจำเป็นต้องใช้คอมพิวเตอร์ถ้าปัญหามีขอบเขตในการหาคำตอบขนาดใหญ่ 3) วิธี GA สามารถที่จะกำหนดพารามิเตอร์และเลือกผลลัพธ์ที่เหมาะสมจากหลายๆ วิธีได้ 4) วิธี GA ให้ผลลัพธ์ที่แม่นยำกว่าและให้ค่าการพยากรณ์ที่แม่นยำกว่า

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลที่ใช้ในการทำนายปริมาณอินทรีย์วัตถุ โดยเปรียบเทียบประสิทธิภาพของโมเดลที่ได้จาก 3 วิธีการได้แก่ วิธีการถดถอยเชิงเส้นพหุ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม เพื่อให้บรรลุวัตถุประสงค์ดังกล่าว ผู้วิจัยดำเนินการวิจัยโดยมีรายละเอียดเกี่ยวกับ ขั้นตอนการดำเนินงานวิจัย รายละเอียดตัวแปรที่ใช้ในงานวิจัยและขั้นตอนการสร้างโมเดลในการพยากรณ์สำหรับแต่ละวิธี ซึ่งสามารถแจกแจงรายละเอียดได้ดังต่อไปนี้

ขั้นตอนการดำเนินงานวิจัย

ในงานวิจัยนี้สามารถแบ่งขั้นตอนในการศึกษาได้เป็น 5 ขั้นตอนดังต่อไปนี้

ขั้นที่ 1 การเก็บรวบรวมข้อมูลจากเอกสารและแหล่งข้อมูลที่เกี่ยวข้อง ได้แก่ข้อมูลคุณสมบัติทางเคมีของดิน

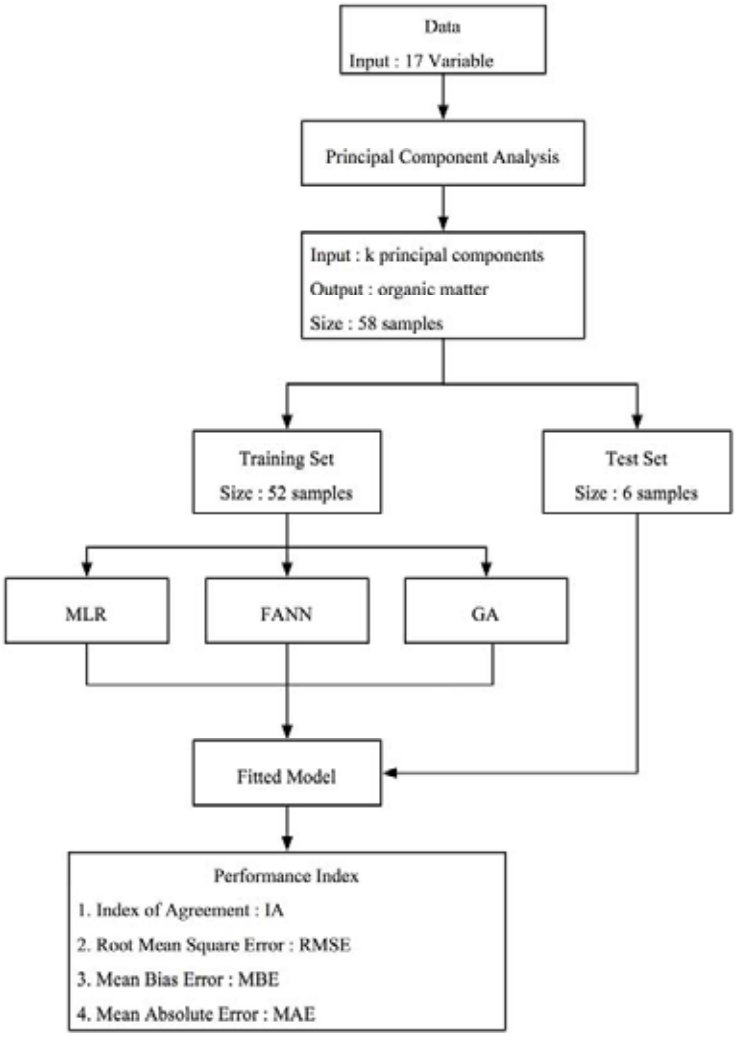
ขั้นที่ 2 วิเคราะห์หาตัวแปรที่เป็นองค์ประกอบหลัก จากตัวแปรคุณสมบัติทางเคมีของดินที่จะใช้เป็นตัวแปรพยากรณ์ในโมเดล ด้วยวิธีวิเคราะห์องค์ประกอบหลัก

ขั้นที่ 3 แบ่งข้อมูลออกเป็นสองส่วนหลัก โดยการสุ่มตัวอย่างอย่างง่ายคือข้อมูลชุดแรก 52 ชุดเป็นชุดสำหรับเรียนรู้ และข้อมูล 6 ชุด เป็นชุดสำหรับการทดสอบโมเดล ในโมเดลจีเนติกอัลกอริทึม สำหรับโมเดลวิธีการถดถอยเชิงเส้นพหุคูณและวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าใช้ข้อมูลชุดทดสอบเดียวกัน แต่ข้อมูลชุดเรียนรู้ 52 ชุด จะถูกแบ่งออกเป็นสองส่วนคือชุดเรียนรู้ 46 ชุด และชุดประเมินผลโมเดล 6 ชุด

ขั้นที่ 4 สร้างโมเดลการพยากรณ์ปริมาณอินทรีย์วัตถุในดินด้วยชุดข้อมูลเรียนรู้ ด้วยวิธีการ 3 วิธีได้แก่วิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม โดยขั้นตอนการสร้างโมเดลจะกล่าวในหัวข้อถัดไป

ขั้นที่ 5 เปรียบเทียบประสิทธิภาพของโมเดลจากทั้ง 3 วิธีในการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน โดยใช้ข้อมูลชุดทดสอบจำนวน 6 ชุด จากการคำนวณค่าวัดประสิทธิภาพ คือ AI, RMSE, MBE และ MAE

แสดงภาพรวมขั้นตอนการศึกษา ได้ดังภาพที่ 11



ภาพที่ 11 ภาพรวมขั้นตอนการศึกษา

ตัวแปรที่ใช้ในการวิจัย

1. พื้นที่ศึกษา คือดินจากพื้นที่เกษตรกรรม ประเภทสวนผลไม้ในภูมิภาคตะวันตก ทั้งหมด 58 ตัวอย่าง ที่เก็บมาจากพื้นที่ 3 จังหวัดได้แก่ จังหวัดนครปฐม จังหวัดสมุทรสาครและจังหวัดสมุทรสงคราม

2. ตัวแปรตาม คือปริมาณอินทรีย์วัตถุ (Organic Matter) เป็นตัวแปรเชิงปริมาณ

3. ตัวแปรอิสระ คือคุณสมบัติทางเคมีของดิน มีจำนวน 17 ตัว ได้แก่ อลูมินัม (Aluminium: Al), แมงกานีส (Manganese: Mn), เหล็ก (Iron: Fe), โครเมียม (Chromium: Cr), แมกนีเซียม (Magnesium: Mg), สังกะสี (Zinc: Zn), ทองแดง (Cupper: Cu), ตะกั่ว (Lead: Pb), โพแทสเซียม (Potassium: K), โซเดียม (Sodium: Na), แคลเซียม (Calcium: Ca), กรดฟัลวิก (Fulvic Acid: FA), กรดฮิวมิก (Humic Acid : HA), ค่าการแลกเปลี่ยนประจุบวก (Cation exchange capacity: CEC), เปอร์เซ็นต์ดินเหนียว (Percent clay: %clay), ไนโตรเจนรวม (TN) และ อินทรีย์คาร์บอน (OC) รายละเอียดของหน่วยวัดของแต่ละตัวแปรมีดังนี้

3.1 อลูมินัม (mg/g soil)

3.2 แมงกานีส (mg/g soil)

3.3 เหล็ก (mg/g soil)

3.4 โครเมียม (mg/g soil)

3.5 แมกนีเซียม (mg/g soil)

3.6 สังกะสี (mg/g soil)

3.7 ทองแดง (mg/g soil)

3.8 ตะกั่ว (mg/g soil)

3.9 โพแทสเซียม (mg/g soil)

3.10 โซเดียม (mg/g soil)

3.11 แคลเซียม (mg/g soil)

3.12 กรดฟัลวิก (mg/g soil)

3.13 กรดฮิวมิก (mg/g soil)

3.14 ค่าการแลกเปลี่ยนประจุบวก (cmol/Kg)

3.15 เปอร์เซ็นต์ดินเหนียว (%)

3.16 ไนโตรเจนรวม (%)

3.17 อินทรีย์คาร์บอน (%)

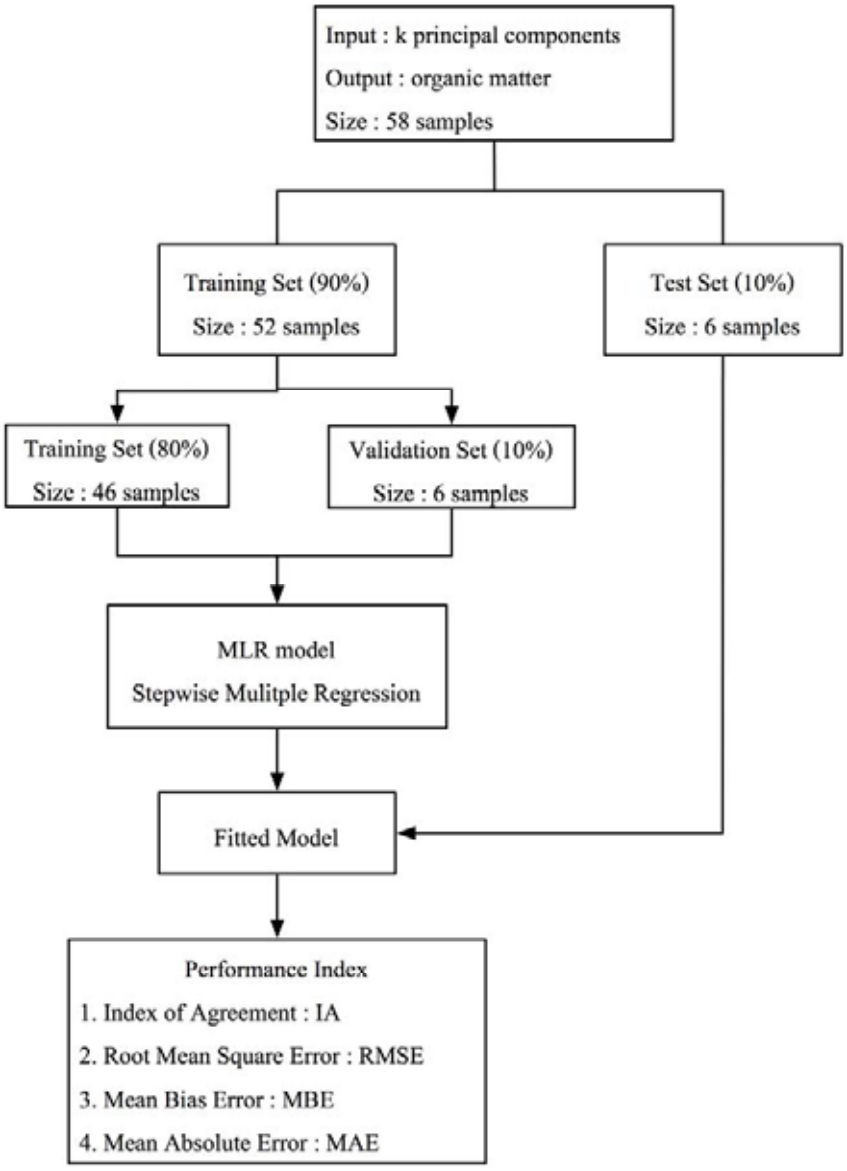
การสร้างโมเดลด้วยวิธีการถดถอยเชิงพหุคูณ

วิธีการถดถอยเชิงพหุคูณเป็นการศึกษาความสัมพันธ์ของตัวแปรตาม (ปริมาณอินทรีย์วัตถุ) และตัวแปรอิสระหลายๆ ตัว (คุณสมบัติทางเคมีของดิน) โดยตัวแปรอิสระที่ศึกษาที่นำมาใช้ในการสร้างโมเดลมาจากวิธีองค์ประกอบหลัก

การคัดเลือกตัวแปรเข้าสู่โมเดล โดยใช้วิธีการถดถอยเชิงพหุคูณแบบเป็นขั้นตอน (Stepwise multiple regression) เป็นวิธีการเลือกตัวแปรอิสระเข้าไปในโมเดลครั้งละ หนึ่งตัวแปร ตัวแปรอิสระที่เข้าไปในโมเดลอาจถูกตัดออกได้ในภายหลัง นั่นคือต้องทดสอบว่าตัวแปรอิสระนั้นเมื่ออยู่ในแบบจำลองแล้ว มีส่วนช่วยในการอธิบายความผันแปรของตัวแปรตาม ขณะที่มีตัวแปรอิสระอื่นๆ อยู่ด้วยหรือไม่

ข้อมูลที่ได้จากการวิเคราะห์องค์ประกอบหลัก 5 องค์ประกอบหลักที่ได้เป็นตัวแปรอิสระ โดยมีปริมาณอินทรีย์วัตถุเป็นตัวแปรตาม แล้วแบ่งข้อมูลโดยการสุ่มอย่างง่าย ออกเป็น 3 ชุด คือ ชุดเรียนรู้ (training set) สำหรับการสร้างโมเดลการถดถอยเชิงเส้นพหุคูณ ชุดประเมินผล (validation set) สำหรับการประเมินผลลัพธ์เพื่อหาโมเดลที่เหมาะสมที่สุด และชุดทดสอบ (test set) สำหรับทดสอบโมเดลที่ได้ โดยแบ่งข้อมูลที่ได้เป็นร้อยละ 80, 10 และ 10 คิดเป็นจำนวนตัวอย่างดิน 46, 6 และ 6 ชุดข้อมูล ตามลำดับ

การตรวจสอบความถูกต้องของโมเดล โดยการหาความสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการพยากรณ์ โดยใช้เกณฑ์ ดัชนีการยอมรับ (IA), ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE), ค่าความคลาดเคลื่อนเอนเอียงเฉลี่ย (MBE) และค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) ตามลำดับ ขั้นตอนการสร้างโมเดลการถดถอยเชิงพหุคูณ แสดงภาพที่ 12



ภาพที่ 12 ขั้นตอนการสร้างโมเดลการถดถอยเชิงเส้นพหุคูณ

การสร้างโมเดลด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า

การเตรียมข้อมูล ข้อมูลที่ได้จากการวิเคราะห์องค์ประกอบหลัก 5 องค์ประกอบหลักที่ได้เป็นตัวแปรอิสระ โดยมีปริมาณอินทรีย์วัตถุเป็นตัวแปรตาม แล้วแบ่งข้อมูลโดยการสุ่มอย่างง่ายออกเป็น 3 ชุด คือ ชุดเรียนรู้ (training set) สำหรับการสร้างโมเดลโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า ชุดประเมินผล (validation set) สำหรับการประเมินผลลัพธ์เพื่อหาโมเดลที่

เหมาะสมที่สุด และชุดทดสอบ (test set) สำหรับทดสอบโมเดลที่ได้ โดยแบ่งข้อมูลที่ได้เป็นร้อยละ 80, 10 และ 10 คิดเป็นจำนวนตัวอย่างดิน 46, 6 และ 6 ชุดข้อมูล ตามลำดับ

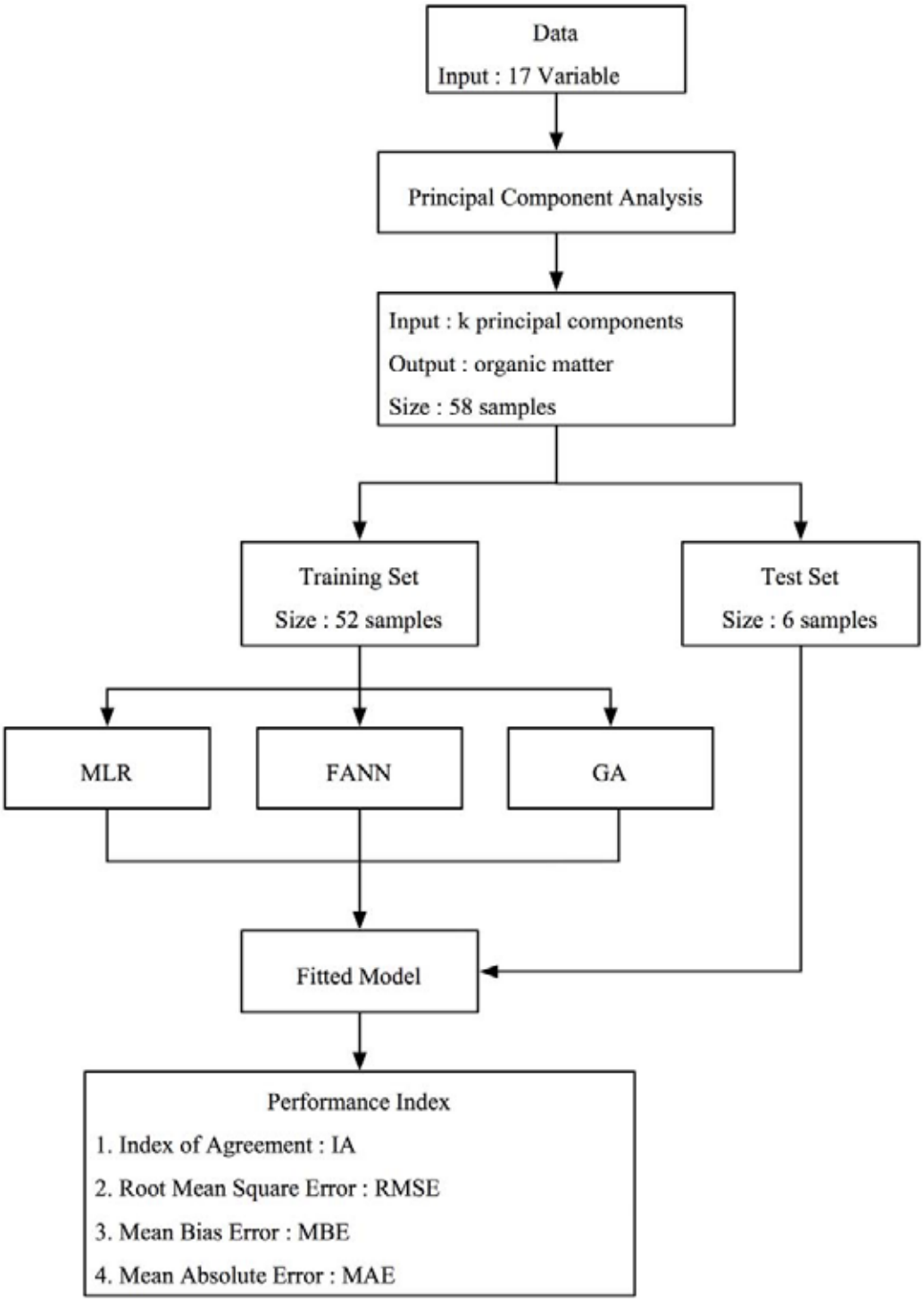
การออกแบบโครงข่าย จะพิจารณาจากจำนวนข้อมูลป้อนเข้าโครงข่าย จำนวนชั้นซ่อน และจำนวนผลลัพธ์เป็นหลัก ซึ่งจะต้องทำการหาค่าให้เหมาะสมกับโครงข่ายมากที่สุด โดยโครงข่ายประสาทเทียมที่ใช้ในการศึกษาครั้งนี้ จะใช้สองตัวแบบ คือ ระบบโครงข่ายประสาทเทียมที่มี 3 ชั้น และ 4 ชั้น และในงานวิจัยนี้จะใช้สถาปัตยกรรมโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า (feed-forward artificial neural network)

การสอนโครงข่าย ในขั้นตอนการสอนโครงข่ายนี้จะใช้ Levenberg-Maquardt back-propagation algorithm โดยใช้ฟังก์ชัน Sigmoid แบบ Tangent Hyperbolic Function เนื่องจากฟังก์ชันการแปลงค่ามีความสำคัญมากในการสอนโครงข่าย ซึ่งฟังก์ชันที่ใช้ควรมีความต่อเนื่อง ไม่เป็นเชิงเส้น ง่ายต่อการคำนวณ โดยฟังก์ชัน Sigmoid แบบ Tangent Hyperbolic Function มีรูปแบบดังสมการที่ 15

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad \text{Range (-1, 1)} \tag{15}$$

ในการใช้ชุดสอน (training set) ให้โครงข่ายเรียนรู้ โครงข่ายจะประมวลผลจนได้คำตอบชุดหนึ่ง สำหรับคำตอบที่โครงข่ายสามารถคำนวณได้ จะถูกนำมาหาความคลาดเคลื่อน โดยเทียบกับปริมาณอินทรีย์วัตถุจริง ถ้ายังมีค่าความคลาดเคลื่อนสูงอยู่ ระบบจะย้อนกลับไปปรับเปลี่ยนค่าถ่วงน้ำหนัก และทำการสอนต่อไป จนกว่าค่าความคลาดเคลื่อนระหว่างตำตอบที่ได้จากโครงข่าย และคำตอบจริง จะมีค่าน้อยในระดับที่ยอมรับได้ ในขณะเดียวกันชุดประเมินผล (validation set) ก็จะทำการทดสอบค่าความคลาดเคลื่อนของโครงข่ายไปพร้อมๆ กัน เมื่อค่าความคลาดเคลื่อนจากชุดประเมินผลมีค่าน้อยกว่าค่าที่ยอมรับได้ จึงหยุดทำการปรับการสอน และได้โครงข่ายที่เหมาะสม

การตรวจสอบความถูกต้องของโครงข่าย หลังจากได้โครงข่ายที่เหมาะสมแล้ว ทำการตรวจสอบความถูกต้องของโมเดล โดยการหาความสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการพยากรณ์ โดยใช้เกณฑ์ ดัชนีการยอมรับ (IA), ค่ารากที่สองของค่าความผิดพลาดกำลังสองเฉลี่ย (RMSE), ค่าความคลาดเคลื่อนเอนเอียงเฉลี่ย (MBE) และค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) ตามลำดับ ขั้นตอนการสร้างโมเดลโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า แสดงดังภาพที่ 13



ภาพที่ 13 ขั้นตอนการสร้างโมเดล FANN

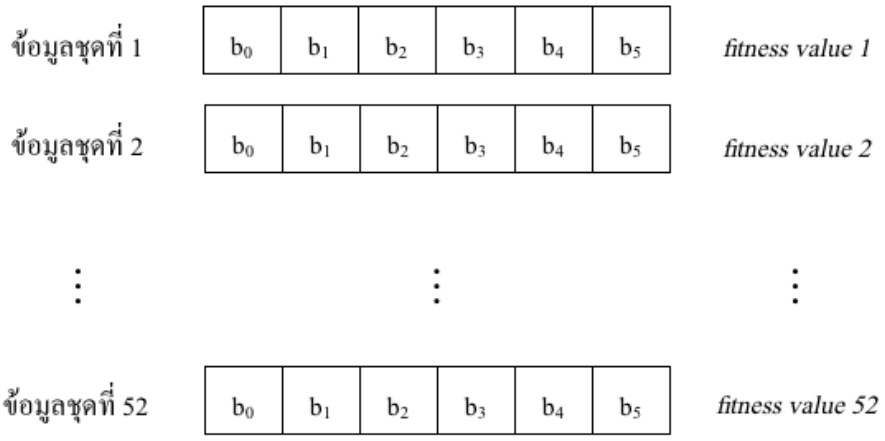
ที่มา : ปรับปรุงจาก Jiang Dahe et al., "Progress in developing an ANN model for air pollution index forecast," Atmospheric Environment 38, 40 (December 2004) : 7055-7064.

การสร้างโมเดลด้วยวิธีจีเนติกอัลกอริทึม

ในงานวิจัยนี้จะใช้วิธีจีเนติกอัลกอริทึมร่วมกับวิธีการถดถอยเชิงเส้นพหุคูณ โดยจะใช้วิธีจีเนติกอัลกอริทึมในการเลือกสัมประสิทธิ์การถดถอยที่ดีที่สุด ให้กับสมการถดถอยเชิงเส้นพหุคูณ ซึ่งสามารถอธิบายได้ตามองค์ประกอบที่สำคัญของจีเนติกอัลกอริทึม มีรายละเอียดดังต่อไปนี้

ประชากรต้นแบบ (Initial Population) คือการสุ่มเลือกเพื่อสร้างประชากรต้นแบบขึ้นมาเพื่อใช้เป็นจุดเริ่มต้นของขั้นตอนการวิวัฒนาการ ในงานวิจัยนี้กำหนดประชากรเท่ากับ 52

การเข้ารหัสโครโมโซม (Chromosome encoding) เป็นขั้นตอนสำหรับแปลงทางเลือกสำหรับการแก้ปัญหาที่เป็นไปได้ให้อยู่ในรูปแบบของโครโมโซม (Chromosome) ซึ่งในงานวิจัยนี้จะใช้ value encoding ในการเข้ารหัสโครโมโซมเนื่องจากค่าที่เราต้องการเข้ารหัสเป็นจำนวนจริง ซึ่งในงานวิจัยจะกำหนดประชากรเท่ากับ 52 ดังนั้นโครโมโซมจะมีเท่ากับ 52 โครโมโซมและแต่ละโครโมโซมจะมีค่าความเหมาะสม (fitness value) ของแต่ละโครโมโซม รูปแบบการสร้างโครโมโซมดังแสดงในภาพที่ 14



ภาพที่ 14 รูปแบบการสร้างโครโมโซม

ฟังก์ชั่นวัตถุประสงค์ (Objective function) เป็นกระบวนการสำคัญที่ใช้ในการประเมินผลคำตอบของระบบว่าดีหรือไม่แค่ไหน ฟังก์ชั่นวัตถุประสงค์เป็นฟังก์ชั่นที่ทำการประเมินคำตอบจากโครโมโซมเทียบกับเป้าหมายของระบบ โดยในงานวิจัยนี้ต้องการค้นหาค่าต่ำสุด

(Minimization) โดยใช้ฟังก์ชั่นหาค่าความเหมาะสม (Fitness Function) คือค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ในการกำหนดค่าความเหมาะสมให้กับแต่ละโครโมโซม เพื่อใช้สำหรับตัดสินใจในการเลือกโครโมโซมเพื่อสืบพันธุ์ในรุ่นถัดไป โดยมีรูปแบบดังต่อไปนี้

$$RMSE = \min \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}$$

โดยที่ Y_i คือค่าจริงของปริมาณอินทรีย์วัตถุในดิน โดยที่ i = 1,2,…,n

$\hat{Y}_i$ คือค่าพยากรณ์ของปริมาณอินทรีย์วัตถุในดิน โดยที่ i = 1,2,…,n

n คือจำนวนตัวอย่างที่นำมาศึกษา

ปฏิบัติการทางสายพันธุ์ (Genetic operation) คือขั้นตอนการดำเนินการต่างๆ ของ GA เพื่อให้การเกิดวิวัฒนาการไปสู่คำตอบที่ดีขึ้น ซึ่งได้แก่

การคัดเลือกสายพันธุ์ (Selection) ซึ่งในงานวิจัยนี้เลือกใช้วิธีการคัดเลือกแบบวิธีการแข่งขัน (Tournament Selection) โดยการสุ่มแบ่งกลุ่มคัดเลือกโครโมโซมแล้วเลือกเอาโครโมโซมที่ดีที่สุดในกลุ่มนั้นเพื่อเป็นต้นกำเนิดสายพันธุ์ วิธีการจัดการแข่งขันมีความเหมาะสมในการทำให้ปัญหาความเหลื่อมล้ำของค่าความเหมาะสมของโครโมโซมหมดไป

ครอสโอเวอร์ (Crossover) เป็นวิธีการรวมตัวใหม่ของโครโมโซม โดยทำการรวมส่วนย่อยระหว่างโครโมโซมต้นกำเนิดสายพันธุ์ตั้งแต่สองโครโมโซมขึ้นไป เพื่อให้กลายเป็นโครโมโซมลูกหลาน โครโมโซมลูกหลานที่ได้จากการครอสโอเวอร์นี้จะมีพันธุกรรมจากต้นกำเนิดสายพันธุ์อยู่ในตัว โดยปกติทั่วไปแล้วจะมีการกำหนดอัตราการทำครอสโอเวอร์เอาไว้ ซึ่งส่วนใหญ่จะใช้ความน่าจะเป็นของการครอสโอเวอร์ (P_c) เป็นตัวกำหนดอัตราการเกิดครอสโอเวอร์ ซึ่งในงานวิจัยนี้จะใช้ความน่าจะเป็นของการครอสโอเวอร์เท่ากับ 0.85 และกำหนดให้มีการทำครอสโอเวอร์แบบจุดเดียว (Single-Point Crossover)

มิวเทชั่น (Mutation) เป็นวิธีการแปรผันยีนหรือส่วนย่อยของโครโมโซม ทำให้เกิดความหลากหลายขึ้นในกลุ่มประชากร มีผลให้คำตอบที่เกิดขึ้นในขบวนการของ GA ครอบคลุมพื้นที่การค้นหาคำตอบทั่วถึงยิ่งขึ้น ซึ่งส่วนใหญ่จะใช้ความน่าจะเป็นของการมิวเทชั่น (P_m) เป็นตัวกำหนดอัตราการเกิดมิวเทชั่นซึ่งในงานวิจัยนี้จะไม่กำหนดความน่าจะเป็นของการมิวเทชั่นเนื่องจากขอบเขตของสัมประสิทธิ์การถดถอยแตกต่างกัน

พารามิเตอร์ (Parameters) การตั้งค่าพารามิเตอร์สำหรับการทำงานของของจีเนติกอัลกอริทึม โดยสรุปคือ

Population size	*52*
Probability of Crossover	*0.85*
Probability of Mutation	*0.00*
Replacement	*steady state*
Selection	*tournament (size = 10)*
Maximum Generation	*100*

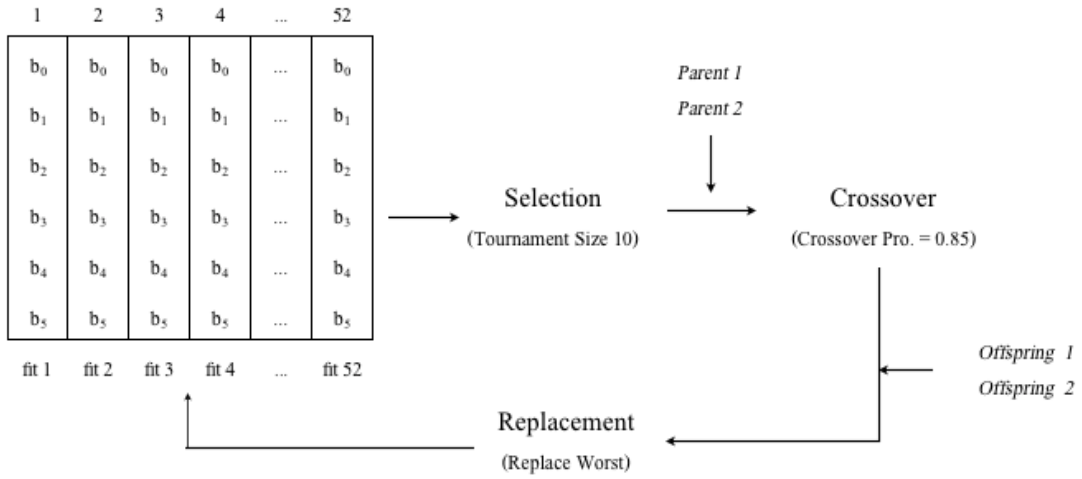
ขอบเขตของการสุ่มสัมประสิทธิ์ (Bounds) จะใช้ขอบเขตของสัมประสิทธิ์การถดถอยที่ได้จากการใช้ชุดเรียนรู้ 52 ชุดในการสร้างสมการถดถอยแบบบังคับตัวแปรเข้าสมการทุกตัวแปร (Enter multiple regression) ในการสร้างช่วงความเชื่อมั่น 95% ของค่าสัมประสิทธิ์การถดถอย และนำมาใช้กำหนดขอบเขตของการสุ่มค่าสัมประสิทธิ์การถดถอยด้วยวิธีจีเนติกอัลกอริทึม ซึ่งมีค่าต่ำสุดและสูงสุดดังตารางที่ 1

ตารางที่ 1 ค่าต่ำสุดและค่าสูงสุดของสัมประสิทธิ์การถดถอยที่ได้จากการสร้างช่วงความเชื่อมั่น 95%

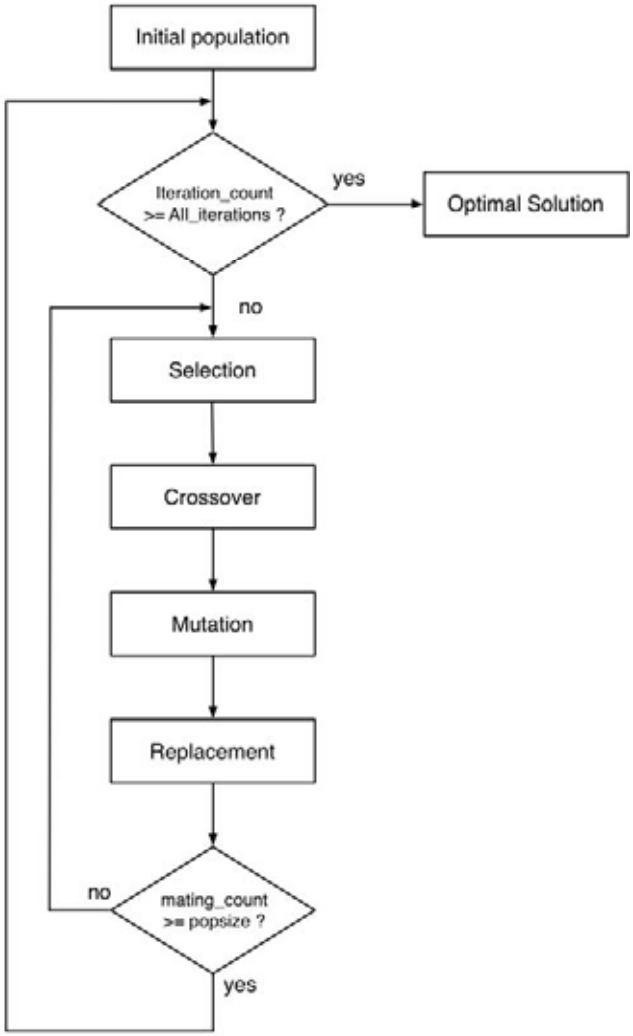
สัมประสิทธิ์การถดถอย	ค่าต่ำสุด	ค่าสูงสุด
b_0	2.376	2.496
b_1	0.104	0.159
b_2	0.319	0.391
b_3	-0.091	-0.008
b_4	-0.037	0.049
b_5	-0.238	-0.123

กระบวนการทำงานของจีเนติกอัลกอริทึม ประกอบด้วยขั้นตอนแรกคือการสร้างประชากรต้นแบบทั้งหมด 52 ตัว ทำการสุ่มโครโมโซมด้วยวิธีการคัดเลือกแบบแข่งขัน (Tournament) ที่มีขนาดเท่ากับ 10 เพื่อหาโครโมโซมตัวที่ดีที่สุด สำหรับเป็นพ่อและแม่ (Parent) ผ่านกระบวนการ

ครอสโอเวอร์เพื่อให้เกิดความหลากหลายทางวิวัฒนาการ โดยในงานวิจัยนี้กำหนดความน่าจะเป็นในการครอสโอเวอร์เท่ากับ 0.85 และจะไม่เกิดมิวเทชันเนื่องจากขอบเขตในการสุ่มของพารามิเตอร์แต่ละตัวมีช่วงการสุ่มค่าที่แตกต่างกัน ซึ่งหลังจากกระบวนการครอสโอเวอร์จะได้ลูกหลาน (Offspring) 2 โครโมโซม เข้าสู่กระบวนการแทนที่โดยใช้วิธีแทนที่แบบคงตัว โดยใช้วิธี Negative Tournament ที่มีขนาดเท่ากับ 10 เพื่อหาโครโมโซมตัวที่แย่ที่สุด แล้วแทนที่ด้วยรุ่นลูกหลานต่อไป กระบวนการนี้จะเกิดขึ้นจนกระทั่งครบจำนวนประชากรที่ตั้งไว้คือ 52 แล้วจึงเริ่มต้นกระบวนการใหม่ ซึ่งแต่ละรอบเราจะเรียกว่าเจนเนอเรชัน (Generation) ทำซ้ำจนกระทั่งครบจำนวนเจเนอเรชันที่กำหนดไว้ ซึ่งในงานวิจัยนี้กำหนดจำนวนเจเนอเรชันเท่ากับ 100 กระบวนการทำงานของจีเนติกอัลกอริทึมที่มีการแทนที่แบบคงตัวในงานวิจัยนี้ แสดงดังรูปที่ 15 โดยกระบวนการทำงานของจีเนติกอัลกอริทึมสามารถแสดงได้ในภาพที่ 16



ภาพที่ 15 กระบวนการทำงานของจีเนติกอัลกอริทึมที่มีการแทนที่แบบคงตัว



ภาพที่ 16 ขั้นตอนการทำงานของจีเนติกอัลกอริทึม

ที่มา : ปรับปรุงจาก Kucuk Ilker and Naim Derebasi, "Prediction of power losses in transformer cores using feed forward neural network and genetic algorithm," Measurement 39, 7 (August 2006) : 607.

เครื่องมือที่ใช้ในการวิเคราะห์

ในงานวิจัยนี้ การวิเคราะห์ตัวแบบการถดถอยเชิงเส้นพหุคูณและโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า ใช้โปรแกรมสำเร็จรูป SAS (Statistical Analysis System) และวิธีจีเนติกอัลกอริทึม เขียนโปรแกรมวิเคราะห์โดยใช้ภาษา JAVA

บทที่ 4
ผลการวิจัย

ในบทนี้ จะนำเสนอผลการวิจัยเป็น 2 ส่วน ส่วนแรกเป็นการวิเคราะห์ข้อมูลคุณสมบัติทางเคมีของดิน และการวิเคราะห์องค์ประกอบหลัก ส่วนที่สองเป็นผลลัพธ์ที่ได้จากการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน โดยประกอบด้วย 2 ตอนหลักคือ (1) การสร้างโมเดลและทดสอบโมเดลจากวิธีการถดถอยเชิงเส้นพหุคูณ, วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม (2) การเปรียบเทียบประสิทธิภาพของตัวแบบทั้งสาม

ส่วนที่ 1 การวิเคราะห์ข้อมูลคุณสมบัติทางเคมีของดินและการวิเคราะห์องค์ประกอบหลัก

ตัวอย่างดินที่ใช้ในการศึกษา เป็นดินจากพื้นที่เกษตรกรรมประเภทสวนผลไม้ในภูมิภาคตะวันตกของประเทศไทย ประกอบด้วยพื้นที่สามจังหวัดของประเทศไทย ได้แก่ นครปฐม สมุทรสาครและสมุทรสงคราม จำนวน 58 จุด ซึ่งนำมาวัดข้อมูลคุณสมบัติทางเคมีในดิน ซึ่งนำมาเป็นตัวแปรที่ใช้พยากรณ์ 17 ตัวแปร ประกอบด้วย อลูมินัม, แมงกานีส, เหล็ก, โครเมียม, แมกนีเซียม, สังกะสี, ทองแดง, ตะกั่ว, โพแทสเซียม, โซเดียม, แคลเซียม, กรดฟัลวิก และกรดฮิวมิก มีหน่วยการวัดเป็น mg/g soil ค่าการแลกเปลี่ยนประจุบวก มีหน่วยการวัดเป็น cmol/Kg ส่วน เปอร์เซ็นต์ดินเหนียว, ไนโตรเจนรวม และอินทรีย์คาร์บอน มีหน่วยการวัดเป็นเปอร์เซ็นต์ (%) ค่าสถิติพื้นฐานของคุณสมบัติทางเคมีของตัวอย่างดินที่นำมาศึกษา ซึ่งจะพบว่าคุณสมบัติทางเคมีของดินมีการกระจายของข้อมูลสูงมาก ยกตัวอย่างเช่น อลูมินัม มีค่าต่ำสุดเท่ากับ 7929.24 ค่าสูงสุดเท่ากับ 34947.96 มีค่าเฉลี่ยเท่ากับ 14163.40 และส่วนเบี่ยงเบนมาตรฐาน 4137.62 รองลงมา เหล็ก มีค่าต่ำสุดเท่ากับ 11810.52 ค่าสูงสุดเท่ากับ 28602.88 มีค่าเฉลี่ยเท่ากับ 19846.78 และส่วนเบี่ยงเบนมาตรฐาน 3394.37 จากข้อมูลทั้งหมดโดยภาพรวมจะพบว่าข้อมูลที่ถูกวัดมามีการกระจายของข้อมูลสูง ค่าสถิติพื้นฐานของคุณสมบัติทางเคมีของตัวอย่างดินที่นำมาศึกษา แสดงดังตารางที่ 2

ตารางที่ 2 ค่าสถิติพื้นฐานของคุณสมบัติทางเคมีของตัวอย่างดินที่นำมาศึกษา

ตัวแปรพยากรณ์ (หน่วยวัด)	ค่าต่ำสุด	ค่าสูงสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน
1) อลูมินัม (mg/g soil)	7929.24	34947.96	14163.40	4137.62
2) แมงกานีส (mg/g soil)	190.15	2180.88	836.92	394.33
3) เหล็ก (mg/g soil)	11810.52	28602.88	19846.78	3394.37
4) โครเมียม (mg/g soil)	13.27	48.71	22.40	4.98
5) แมกนีเซียม (mg/g soil)	1270.10	5717.70	3131.85	820.79
6) สังกะสี (mg/g soil)	31.36	114.10	60.78	19.91
7) ทองแดง (mg/g soil)	11.47	267.44	39.49	42.46
8) ตะกั่ว (mg/g soil)	24.01	61.38	37.83	8.33
9) โพแทสเซียม (mg/g soil)	747.57	4831.45	2167.00	746.72
10) โซเดียม (mg/g soil)	437.32	5101.70	1697.41	643.47
11) แคลเซียม (mg/g soil)	3957.08	16816.74	12475.71	2567.58
12) กรดฟัลวิก (mg/g soil)	10.65	609.80	151.60	113.61
13) กรดฮิวมิก (mg/g soil)	6.57	49.47	32.91	12.24
14) ค่าการแลกเปลี่ยนประจุบวก (CEC) (cmol/Kg)	14.73	34.29	24.68	4.79
15) เปอร์เซ็นต์ดินเหนียว (%)	14.27	87.87	44.28	15.21
16) ไนโตรเจนรวม (%)	0.07	0.22	0.14	0.04
17) อินทรีย์คาร์บอน (%)	0.70	2.32	1.41	0.42

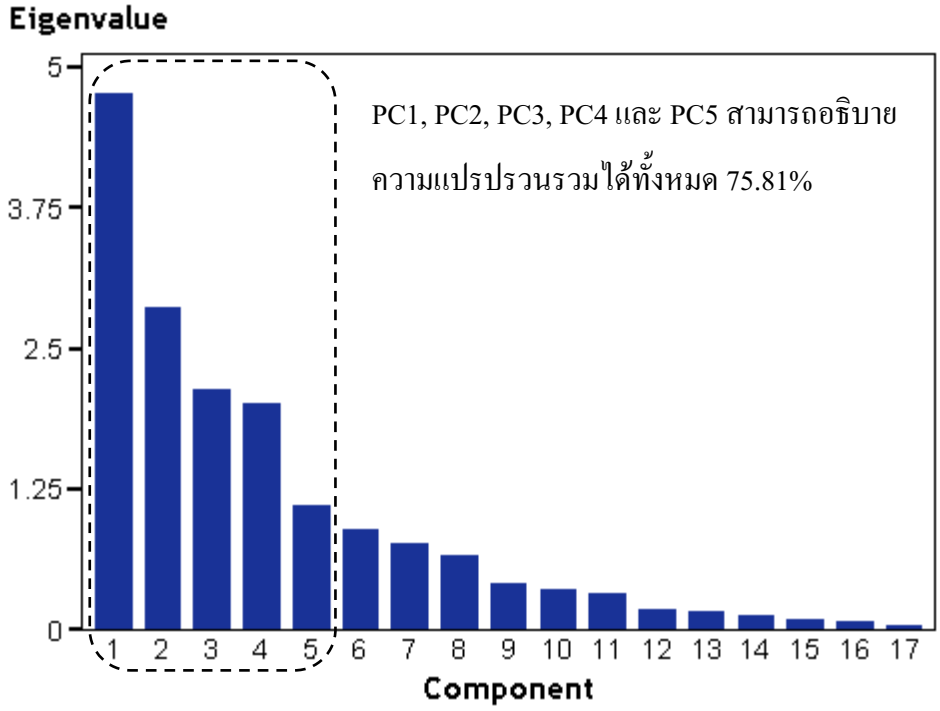
เนื่องจากในการสร้างโมเดลให้มีประสิทธิภาพ กระบวนการประมวลข้อมูลขั้นต้น (Data preprocessing) มีความสำคัญมากเนื่องจากการมีข้อมูลที่ไม่จำเป็นในโมเดลน้อยยิ่งเป็นผลดี ซึ่งงานวิจัยนี้ใช้วิธีวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) เป็นเครื่องมือในการลดจำนวนมิติของข้อมูลและขจัดความสัมพันธ์ระหว่างตัวแปร เพื่อลดความซับซ้อนของโมเดล และทำให้โมเดลสามารถที่จะอธิบายความสัมพันธ์จริงของข้อมูลได้มากยิ่งขึ้น ซึ่งความเหมาะสมของการใช้ PCA บนชุดข้อมูลที่ใช้ในการศึกษานี้ถูกวัดโดยใช้การทดสอบ Bartlett's sphericity โดยสมมติฐานหลักที่เราพิจารณาได้แก่

H_0 : *ตัวแปรทุกตัวไม่มีสหสัมพันธ์ต่อกัน*

H_1 : *มีตัวแปรบางตัวที่มีสหสัมพันธ์ต่อกัน*

จากผลลัพธ์การทดสอบสมมติฐานพบว่าค่า Bartlett's sphericity มีค่าเท่ากับ 681.672 (p-value = .000) ดังนั้นปฏิเสธสมมติฐานหลัก นั่นคือตัวแปรมีสหสัมพันธ์ต่อกัน ดังนั้น PCA ควรถูกนำมาใช้

การเลือกค่าไอเกน โดยเลือกองค์ประกอบที่มีค่าไอเกนมากกว่า 1 ซึ่งมีเพียง 5 องค์ประกอบ เมื่อพิจารณาความแปรปรวนรวมที่ถูกอธิบายโดยองค์ประกอบหลักทั้งหมด พบว่าองค์ประกอบหลักที่ 1 (PC1) จนถึงองค์ประกอบหลักที่ 5 (PC5) สามารถอธิบายความแปรปรวนรวมได้ทั้งหมด 75.81% แต่องค์ประกอบหลักที่ 6 (PC6) ถึงองค์ประกอบหลักที่ 17 (PC17) อธิบายความแปรปรวนได้เพียง 24.19% ดังนั้นจึงเลือก 5 องค์ประกอบหลักแรกที่ได้ มาใช้เป็นตัวแปรพยากรณ์ปริมาณอินทรีย์วัตถุในดินในโมเดลการถดถอยเชิงเส้นพหุคูณ โครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าและจีเนติกอัลกอริทึม การเลือกองค์ประกอบหลัก ดังแสดงในภาพที่ 17



ภาพที่ 17 Scree plot สำหรับการเลือกองค์ประกอบหลัก

ในการเลือกค่าไอเกน (eigenvalues) พบว่ามี 5 องค์ประกอบหลักที่ได้จากวิธี PCA โดยในงานวิจัยนี้เลือกตัวแปรที่มีค่าน้ำหนักมากกว่า 0.40 ขึ้นไป ซึ่งพบว่าองค์ประกอบหลักที่ 1 ประกอบด้วย อลูมินัม, เหล็ก และโครเมียม องค์ประกอบหลักที่ 2 ประกอบด้วย สังกะสี, แคลเซียม, ไนโตรเจนรวม และอินทรีย์คาร์บอน องค์ประกอบหลักที่ 3 ประกอบด้วย ทองแดง, ตะกั่ว, ค่าการแลกเปลี่ยนประจุบวก และเปอร์เซ็นต์ดินเหนียว องค์ประกอบหลักที่ 4 ประกอบด้วย แมงกานีส, แมกนีเซียม และโซเดียม องค์ประกอบหลักที่ 5 ประกอบด้วย โพแทสเซียม, กรดฟัลวิก และกรดฮิวมิก ค่าน้ำหนักของ 5 องค์ประกอบหลักแรกที่ได้ แสดงในตารางที่ 3

ตารางที่ 3 ค่าน้ำหนักของ 5 องค์ประกอบหลักแรกของคุณสมบัติทางเคมีของดิน

องค์ประกอบ	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
อลูมินัม (Al)	**0.946**	0.042	0.090	0.022	0.014
แมงกานีส (Mn)	-0.076	-0.071	0.236	**0.788**	-0.114
เหล็ก (Fe)	**0.775**	0.017	-0.339	0.316	0.012
โครเมียม (Cr)	**0.937**	-0.058	-0.086	0.007	0.158
แมกนีเซียม (Mg)	0.109	0.013	-0.166	**0.874**	0.179
สังกะสี (Zn)	0.535	**0.649**	-0.020	-0.061	0.006
ทองแดง (Cu)	-0.042	0.268	**0.547**	-0.278	0.271
ตะกั่ว (Pb)	0.442	0.109	**-0.669**	0.206	0.208
โพแทสเซียม (K)	0.377	0.191	-0.299	0.306	**0.611**
โซเดียม (Na)	0.291	0.009	-0.071	**0.672**	0.293
แคลเซียม (Ca)	-0.240	**0.653**	0.084	0.323	-0.054
กรดฟัลวิก (FA)	0.022	0.310	-0.226	0.016	**0.812**
กรดฮิวมิก (HA)	0.028	0.039	-0.042	0.104	**0.794**
ค่าการแลกเปลี่ยนประจุบวก (CEC)	-0.125	0.022	**0.775**	0.091	-0.302
เปอร์เซ็นต์ดินเหนียว (% clay)	0.093	-0.037	**0.858**	0.157	-0.131
ไนโตรเจนรวม (TN)	0.117	**0.879**	0.014	-0.142	0.324
อินทรีย์คาร์บอน (OC)	-0.006	**0.931**	-0.024	-0.092	0.190
ค่าไอเกน	4.766	2.868	2.135	2.009	1.108
เปอร์เซ็นต์ความแปรปรวนสะสม	0.280	0.449	0.575	0.693	0.758

ส่วนที่ 2 ผลลัพธ์ที่ได้จากการพยากรณ์ปริมาณอินทรีย์วัตถุในดิน

ประกอบด้วย 2 ตอนหลักคือ (1) ผลลัพธ์ของการพยากรณ์ปริมาณอินทรีย์วัตถุจากวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม (2) การเปรียบเทียบประสิทธิภาพของการพยากรณ์จากวิธีทั้งสามวิธี ซึ่งมีรายละเอียดดังต่อไปนี้

ในการสร้างโมเดลและทดสอบโมเดล โดยใช้วิธีวิธีการถดถอยเชิงเส้นพหุคูณ, วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า จะใช้ชุดข้อมูลทั้งหมด 58 ชุด ซึ่งแบ่งออกเป็นสองชุดอย่างสุ่ม คือชุดข้อมูลสำหรับการเรียนรู้ (training set) จำนวน 52 ชุด และชุดข้อมูลทดสอบ (test set) จำนวน 6 ชุด สำหรับข้อมูลชุดทดสอบโมเดลได้แก่ ข้อมูลชุดที่ 29, 33, 34, 48, 50 และ 52 ตามลำดับ ที่เหลือเป็นชุดเรียนรู้ ส่วนวิธีจีเนติกอัลกอริทึม ใช้เพียงชุดทดสอบ ชุดเดียวกันกับทั้งสองโมเดล

ผลลัพธ์จากวิธีการถดถอยเชิงเส้นพหุคูณ

จากการสร้างโมเดล MLR โดยใช้วิธีการถดถอยพหุคูณแบบเป็นขั้นตอน (Stepwise multiple regression) สมการถดถอยสามารถอธิบายความผันแปรของปริมาณอินทรีย์วัตถุ (R^2) ได้ 89% สามารถแสดงสมการถดถอยเชิงเส้นพหุคูณได้ดังสมการที่ 16 และค่าประมาณพารามิเตอร์ที่ได้จากวิธีการถดถอยเชิงเส้นพหุคูณ แสดงในตารางที่ 4

$$\hat{Y} = 2.423 + 0.135PC_1 + 0.367PC_2 - 0.184PC_5 \quad (16)$$

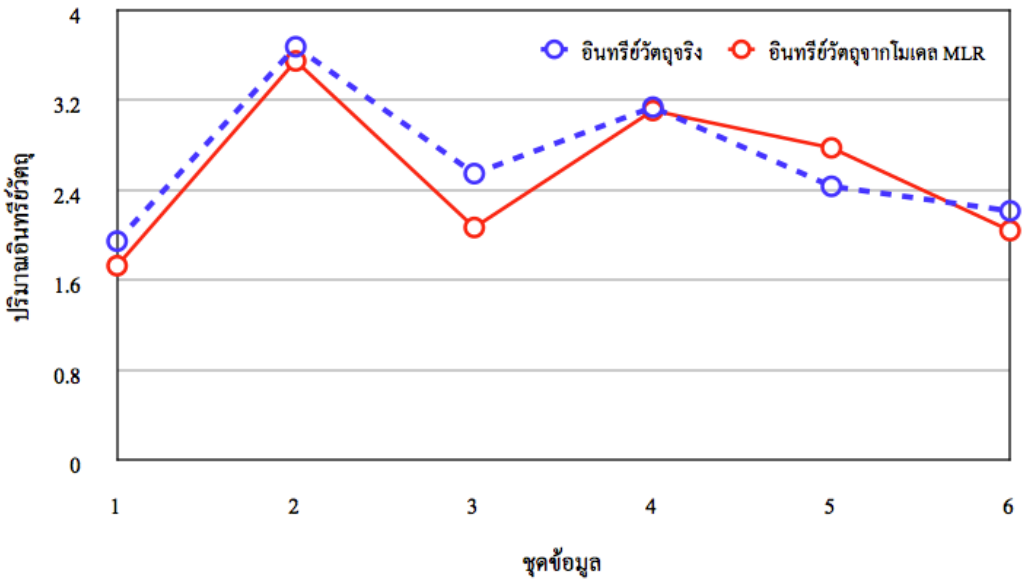
ตารางที่ 4 ค่าประมาณของพารามิเตอร์จากวิธีการถดถอยเชิงเส้นพหุคูณ

Model	Constant	PC1	PC2	PC5
Parameter estimate	2.423	0.135	0.367	-0.184
Standard error	0.034	0.015	0.020	0.030

ผลลัพธ์จากการพยากรณ์โดยใช้ชุดทดสอบ พบว่าปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยโมเดลการถดถอยเชิงเส้นพหุคูณเมื่อเปรียบเทียบกับปริมาณอินทรีย์วัตถุจริง ซึ่งจะเห็นได้ว่าปริมาณอินทรีย์วัตถุจริงกับปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ค่อนข้างใกล้เคียงกัน แสดงได้ดังตารางที่ 5 และเมื่อพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์กับปริมาณอินทรีย์วัตถุจริง ได้ดังภาพที่ 18

ตารางที่ 5 ปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ

ข้อมูล	ปริมาณอินทรีย์วัตถุจริง	ปริมาณอินทรีย์วัตถุจากการพยากรณ์
1	1.9477	1.7296
2	3.6756	3.5493
3	2.5483	2.0695
4	3.1388	3.1084
5	2.4351	2.7770
6	2.2166	2.0418



ภาพที่ 18 การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี MLR ของชุดข้อมูลทดสอบ

เมื่อพิจารณาค่าวัดประสิทธิภาพของโมเดลโดยใช้ ดัชนีการยอมรับ (IA) เป็นค่าที่แสดงความถูกต้องของโมเดลค่าวัดที่ดีที่สุดของค่าดัชนีการยอมรับ มีค่าเท่ากับ 1 ซึ่งจากงานวิจัยนี้พบว่ามีค่าเท่ากับ 0.97 ซึ่งพบว่าค่าวัดประสิทธิภาพที่ได้มีค่าสูงมาก ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE), ค่าความคลาดเคลื่อนเอนเอียงเฉลี่ย (MBE) และค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) ค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ซึ่งจากงานวิจัยนี้พบว่าค่าวัดประสิทธิภาพมีค่าเท่ากับ 0.27, -0.11 และ 0.23 ตามลำดับ ซึ่งพบว่าค่าวัดประสิทธิภาพที่ได้มีค่าน้อย ซึ่งแสดงถึงความแม่นยำในการพยากรณ์สูง ซึ่งเมื่อพิจารณาผลลัพธ์โดยภาพรวมพบว่าค่าวัดประสิทธิภาพที่ได้ค่อนข้างมีความแม่นยำ ประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ แสดงดังตารางที่ 6

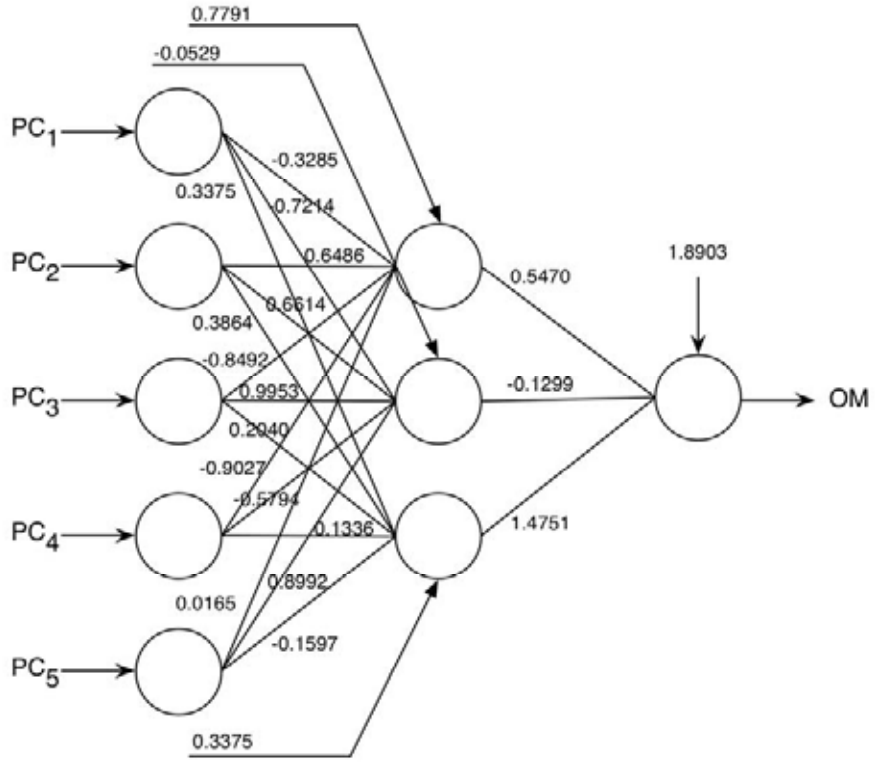
ตารางที่ 6 ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ

ประสิทธิภาพ	วิธี MLR
AI	0.97
RMSE	0.27
MBE	-0.11
MAE	0.23

ผลลัพธ์จากวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า

จากการใช้วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า ในการเรียนรู้และปรับน้ำหนักโครงข่าย ซึ่งพบว่าโมเดลที่ดีที่สุดประกอบด้วย จำนวนโหนดในชั้นอินพุต เท่ากับ 5 จำนวนโหนดในชั้นซ่อน เท่ากับ 3 และจำนวนโหนดในชั้นเอาท์พุต เท่ากับ 1 โดยมีค่าเอนเอียงเท่ากับจำนวนโหนดในชั้นซ่อนรวมกับโหนดในชั้นผลลัพธ์ ซึ่งเท่ากับ 4 แสดงค่าน้ำหนักและค่าเอนเอียงของแต่ละโหนด ได้ดังภาพที่ 19

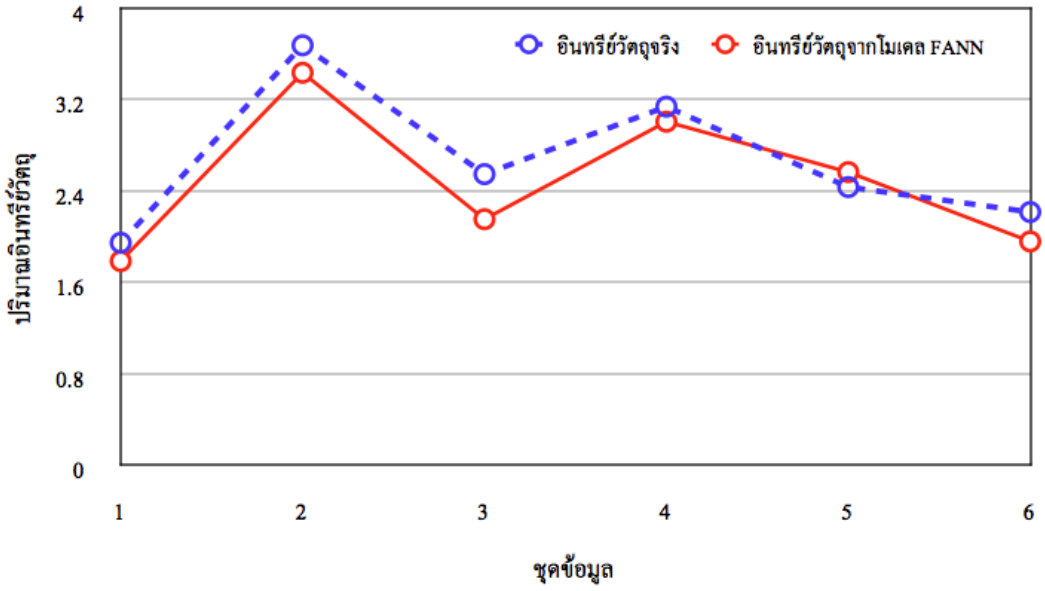
ผลลัพธ์จากการพยากรณ์โดยใช้ชุดทดสอบ พบว่าปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า เมื่อเปรียบเทียบกับปริมาณอินทรีย์วัตถุจริงค่อนข้างใกล้เคียงกัน แสดงดังตารางที่ 7 และเมื่อพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์กับปริมาณอินทรีย์วัตถุจริง ได้ดังภาพที่ 20



ภาพที่ 19 ค่าน้ำหนักของโมเดลที่ดีที่สุดของการเรียนรู้โครงข่ายด้วยวิธี FANN

ตารางที่ 7 ปริมาณอินทรีย์วัตถุจากการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า

ข้อมูล	ปริมาณอินทรีย์วัตถุจริง	ปริมาณอินทรีย์วัตถุจากการพยากรณ์
1	1.9477	1.7872
2	3.6756	3.4354
3	2.5483	2.1539
4	3.1388	3.0072
5	2.4351	2.5659
6	2.2166	1.9595



ภาพที่ 20 การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี FANN ของชุดข้อมูลทดสอบ

เมื่อพิจารณาค่าวัดประสิทธิภาพของโมเดลโดยใช้ ดัชนีการยอมรับ (IA) เป็นค่าที่แสดงความถูกต้องของโมเดลค่าวัดที่ดีที่สุดของค่าดัชนีการยอมรับ มีค่าเท่ากับ 1 ซึ่งจากงานวิจัยนี้พบว่ามีค่าเท่ากับ 0.99 ซึ่งพบว่าค่าวัดประสิทธิภาพที่ได้มีค่าสูงมาก ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE), ค่าความคลาดเคลื่อนเอนเอียงเฉลี่ย (MBE) และค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) ค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ซึ่งจากงานวิจัยนี้พบว่าค่าวัดประสิทธิภาพมีค่าเท่ากับ 0.24, -0.17 และ 0.22 ตามลำดับ ซึ่งพบว่าค่าวัดประสิทธิภาพที่ได้มีค่าน้อย ซึ่งแสดงถึงความแม่นยำในการพยากรณ์สูง ซึ่งเมื่อพิจารณาผลลัพธ์โดยภาพรวมพบว่าค่าวัดประสิทธิภาพที่ได้ค่อนข้างมีความแม่นยำสูง ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า แสดงดังตารางที่ 8

ตารางที่ 8 ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า

ประสิทธิภาพ	วิธี FANN
AI	0.99
RMSE	0.24
MBE	-0.17
MAE	0.22

ผลลัพธ์จากวิธีจีเนติกอัลกอริทึม

จากการใช้วิธีจีเนติกอัลกอริทึมในการเลือกสัมประสิทธิ์การถดถอยที่ดีที่สุด ให้กับสมการการถดถอยเชิงเส้นพหุคูณ สามารถแสดงสมการถดถอยเชิงเส้นพหุคูณได้ดังสมการที่ 17

$$\hat{Y} = 2.440 + 0.117PC_1 + 0.367PC_2 - 0.809PC_3 + 0.047PC_4 - 0.160PC_5 \quad (17)$$

ตารางที่ 9 ค่าสัมประสิทธิ์ของสมการการถดถอยเชิงเส้นพหุคูณ

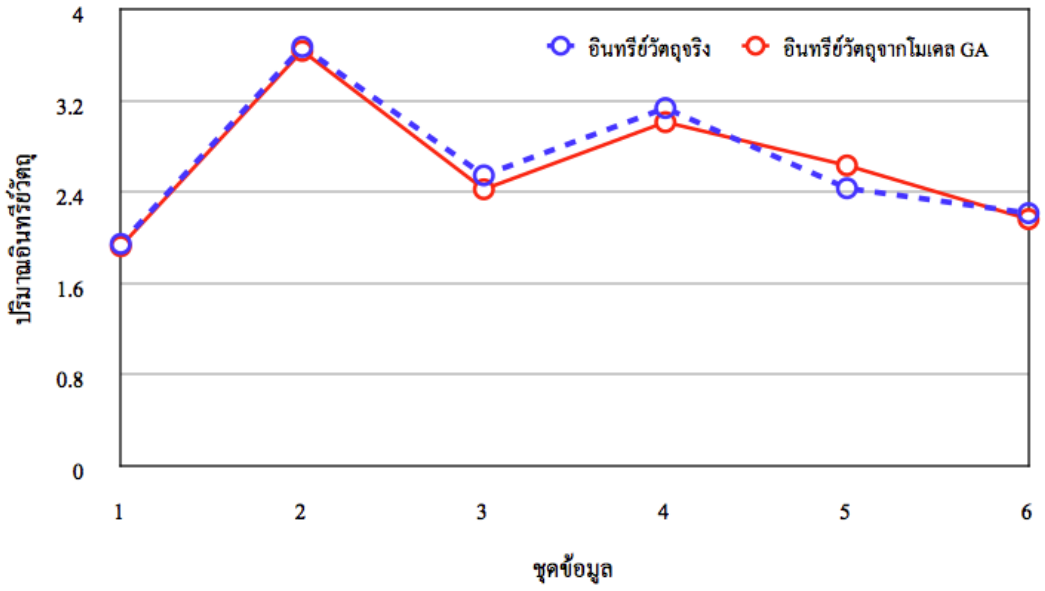
สัมประสิทธิ์การถดถอย	ค่าสัมประสิทธิ์การถดถอย
b_0	2.440
b_1	0.117
b_2	0.367
b_3	-0.089
b_4	0.047
b_5	-0.160

ค่าสัมประสิทธิ์การถดถอยที่ดีที่สุดที่ได้จากวิธีจีเนติกอัลกอริทึม แสดงในตารางที่ 9 ซึ่งเมื่อพิจารณาจะพบว่ามีสัมประสิทธิ์บางตัวเช่น b_3 และ b_4 ที่มีนัยสำคัญน้อยมากสามารถตัดตัวแปรออกจากสมการได้ แต่ในงานวิจัยนี้จะพิจารณาสัมประสิทธิ์ทุกตัวและถือว่าตัวแปรทุกตัวสามารถ

อธิบายความผันแปรของปริมาณอินทรีย์วัตถุได้ ผลลัพธ์จากการพยากรณ์โดยใช้ชุดทดสอบ พบว่าปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม เมื่อเปรียบเทียบกับปริมาณอินทรีย์วัตถุจริงมีความใกล้เคียงกันมาก แสดงดังตารางที่ 10 และเมื่อพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์กับปริมาณอินทรีย์วัตถุจริง แสดงดังภาพที่ 21

ตารางที่ 10 ปริมาณอินทรีย์วัตถุจากการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม

ข้อมูล	ปริมาณอินทรีย์วัตถุจริง	ปริมาณอินทรีย์วัตถุจากการพยากรณ์
1	1.9477	1.9264
2	3.6756	3.6401
3	2.5483	2.4262
4	3.1388	3.0152
5	2.4351	2.6354
6	2.2166	2.1645



ภาพที่ 21 การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี GA ของชุดข้อมูลทดสอบ

เมื่อพิจารณาค่าวัดประสิทธิภาพของโมเดลโดยใช้ ดัชนีการยอมรับ (IA) เป็นค่าที่แสดงความถูกต้องของโมเดลค่าวัดที่ดีที่สุดของค่าดัชนีการยอมรับ มีค่าเท่ากับ 1 ซึ่งจากงานวิจัยนี้พบว่ามีค่าเท่ากับ 0.99 ซึ่งพบว่าค่าวัดประสิทธิภาพที่ได้มีค่าสูงมาก ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE), ค่าความคลาดเคลื่อนเอนเอียงเฉลี่ย (MBE) และค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) ค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ซึ่งจากงานวิจัยนี้พบว่าค่าวัดประสิทธิภาพมีค่าเท่ากับ 0.11, -0.03 และ 0.09 ตามลำดับ ซึ่งพบว่าค่าวัดประสิทธิภาพที่ได้มีค่าน้อย ซึ่งแสดงถึงความแม่นยำในการพยากรณ์สูง ซึ่งเมื่อพิจารณาผลลัพธ์โดยภาพรวมพบว่าค่าวัดประสิทธิภาพที่ได้มีความแม่นยำสูง ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม แสดงดังตารางที่ 11

ตารางที่ 11 ค่าวัดประสิทธิภาพที่ได้จากการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม

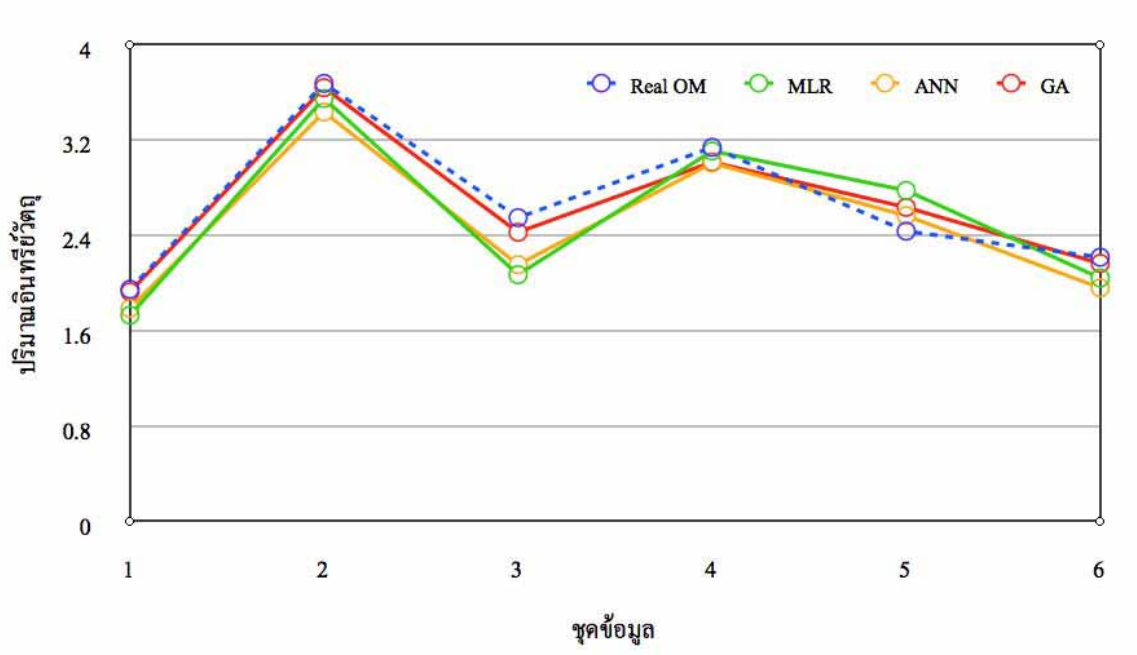
ประสิทธิภาพ	**วิธี GA**
AI	0.99
RMSE	0.11
MBE	-0.03
MAE	0.09

การเปรียบเทียบประสิทธิภาพของการพยากรณ์จากวิธีทั้งสามวิธี

ผลลัพธ์จากการพยากรณ์โดยใช้ชุดทดสอบ 6 ชุด พบว่าปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม เมื่อเปรียบเทียบกับปริมาณอินทรีย์วัตถุจริง วิธีการพยากรณ์ที่ได้จากวิธีจีเนติกอัลกอริทึมให้ค่าที่ใกล้เคียงกับค่าปริมาณอินทรีย์วัตถุจริงมากที่สุด ปริมาณอินทรีย์วัตถุจากการพยากรณ์จากทั้ง 3 โมเดล แสดงดังตารางที่ 12 และพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุที่ได้จากการพยากรณ์กับปริมาณอินทรีย์วัตถุจริง แสดงดังภาพที่ 22

ตารางที่ 12 ปริมาณอินทรีย์วัตถุจากการพยากรณ์จากทั้ง 3 โมเดล

ข้อมูล	ปริมาณอินทรีย์วัตถุจริง	MLR	FANN	GA
1	1.9477	1.7296	1.7872	1.9264
2	3.6756	3.5493	3.4354	3.6401
3	2.5483	2.0695	2.1539	2.4262
4	3.1388	3.1084	3.0072	3.0152
5	2.4351	2.7770	2.5659	2.6354
6	2.2166	2.0418	1.9595	2.1645



ภาพที่ 22 การพล็อตกราฟเปรียบเทียบปริมาณอินทรีย์วัตถุจริงและอินทรีย์วัตถุที่ได้จากการพยากรณ์ด้วยวิธี MLR, FANN และ GA ของชุดข้อมูลทดสอบ

ผลลัพธ์จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม พบว่าการพยากรณ์ที่ได้จากวิธีจีเนติกอัลกอริทึมให้ผลลัพธ์ในการพยากรณ์แม่นยำกว่าการพยากรณ์ที่ได้จากวิธีอื่นๆ พิจารณาค่าวัดประสิทธิภาพ AI ซึ่งค่าวัดที่ดีที่สุดมีค่าเท่ากับ 1 การพยากรณ์ด้วยวิธีการ FANN และ GA ให้ค่า

เท่ากันคือ 0.99 ซึ่งมากกว่าการพยากรณ์ด้วยวิธี MLR ที่มีค่าเท่ากับ 0.97 เมื่อพิจารณาค่าวัดประสิทธิภาพ RMSE และ MAE ซึ่งค่าวัดที่ดีที่สุดมีค่าเท่ากับ 0 ให้ผลลัพธ์สอดคล้องกันคือ การพยากรณ์ด้วยวิธีการ GA ให้ค่าวัดประสิทธิภาพที่ดีที่สุดคือ 0.11 และ 0.09 รองลงมาคือ การพยากรณ์ด้วยวิธีการ FANN ให้ค่าเท่ากับ 0.24 และ 0.22 และการพยากรณ์ด้วยวิธีการ MLR ให้ค่าวัดประสิทธิแย่ที่สุด เท่ากับ 0.27 และ 0.23 ส่วนค่าวัดประสิทธิภาพ MBE วิธีการที่ให้ค่าวัดประสิทธิภาพที่ดีที่สุดคือ การพยากรณ์ด้วยวิธี GA มีค่าเท่ากับ -0.03 รองลงมาการพยากรณ์ด้วยวิธี MLR มีค่าเท่ากับ -0.11 และการพยากรณ์ด้วยวิธี FANN ให้ค่าวัดประสิทธิภาพน้อยที่สุด เท่ากับ -0.17 แสดงค่าวัดประสิทธิภาพจากการพยากรณ์ทั้ง 3 วิธีได้ดังตารางที่ 13

ตารางที่ 13 ค่าวัดประสิทธิภาพการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ (MLR) วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า (FANN) และวิธีจีเนติกอัลกอริทึม (GA)

ประสิทธิภาพ	**MLR**	**FANN**	**GA**
AI	0.97	0.99	0.99
RMSE	0.27	0.24	0.11
MBE	-0.11	-0.17	-0.03
MAE	0.23	0.22	0.09

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การพยากรณ์อินทรีย์วัตถุในดิน จากพื้นที่เกษตรกรรมประเภทสวนผลไม้ในภูมิภาคตะวันตก 3 จังหวัดได้แก่ จังหวัดนครปฐม จังหวัดสมุทรสาครและจังหวัดสมุทรสงคราม โดยเก็บข้อมูลตัวอย่างดินมา 58 ชุด นำมาวิเคราะห์คุณสมบัติทางเคมีของดิน ประกอบด้วย 17 ตัวแปร ผู้วิจัยได้ทำการศึกษา เพื่อให้การพยากรณ์ให้มีความแม่นยำที่สุด ในการสร้างโมเดลให้มีประสิทธิภาพ กระบวนการประมวลข้อมูลขั้นต้น (Data preprocessing) มีความสำคัญมาก เนื่องจากการมีข้อมูลที่ไม่จำเป็นในโมเดลน้อยยิ่งเป็นผลดี ซึ่งงานวิจัยนี้ใช้วิธี PCA เป็นเครื่องมือในการลดจำนวนมิติของข้อมูลและขจัดความสัมพันธ์ระหว่างตัวแปร (Multicollinearity) อิสระ และทำให้โมเดลสามารถที่จะอธิบายความสัมพันธ์จริงของข้อมูลได้มากยิ่งขึ้น แล้วใช้องค์ประกอบหลักใหม่ที่ได้ สร้างโมเดลและทดสอบโมเดลด้วยวิธีการ 3 วิธี ได้แก่ วิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม แล้วเปรียบเทียบผลลัพธ์ที่ได้จากค่าวัดประสิทธิภาพ

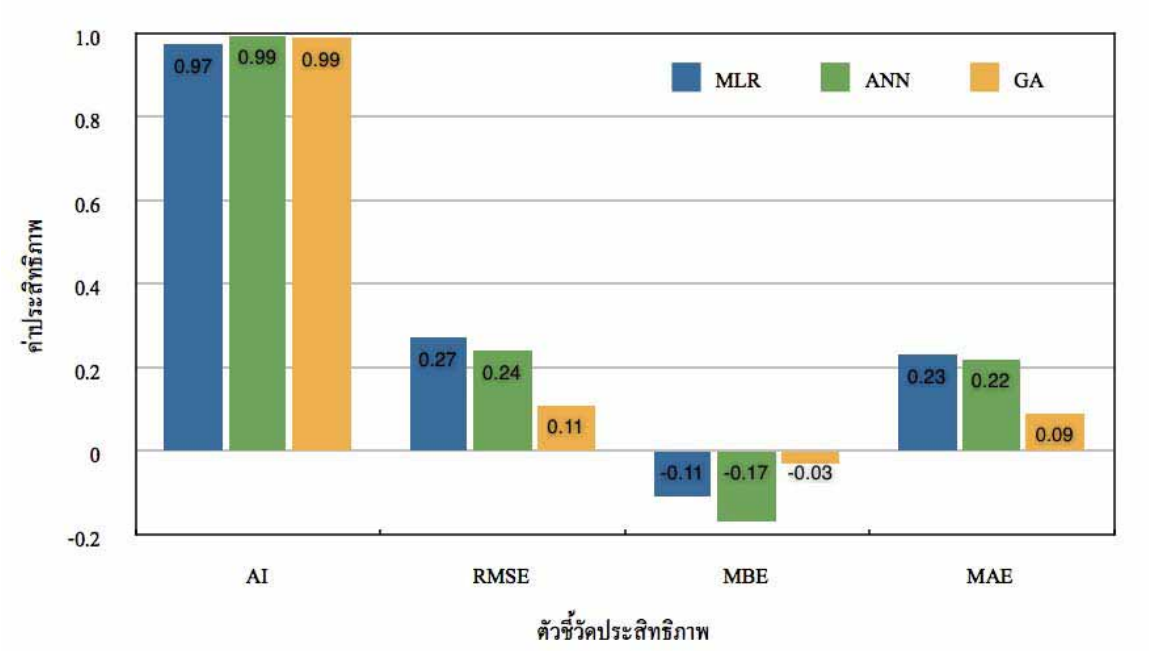
สรุปผลการวิเคราะห์องค์ประกอบหลัก

จากวิธีการวิเคราะห์องค์ประกอบหลัก พบว่าตัวแปรพยากรณ์ทั้ง 17 ตัวแปร ที่ใช้ในการพยากรณ์ปริมาณอินทรีย์วัตถุ เมื่อพิจารณาความสัมพันธ์ระหว่างตัวแปรพบว่ามีตัวแปรบางตัวที่มีความสัมพันธ์กันสูง หลังจากใช้วิธีการวิเคราะห์องค์ประกอบหลัก โดยเลือกองค์ประกอบที่มีค่าไอเกนมากกว่า 1 มีเพียง 5 องค์ประกอบ และสามารถอธิบายความแปรปรวนรวมได้ทั้งหมด 75.81% ซึ่งจะใช้ 5 องค์ประกอบหลักที่ได้ในการพยากรณ์อินทรีย์วัตถุต่อไป

สรุปผลการเปรียบเทียบประสิทธิภาพจากการพยากรณ์

ผลลัพธ์จากการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า และวิธีจีเนติกอัลกอริทึม แสดงค่าวัดประสิทธิภาพการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าและวิธีจีเนติกอัลกอริทึม ซึ่งพบว่าการพยากรณ์ที่ได้จากวิธีจีเนติกอัลกอริทึมให้ผลลัพธ์ในการพยากรณ์แม่นยำกว่าการพยากรณ์ที่ได้จากวิธีอื่นๆ พิจารณาค่าวัดประสิทธิภาพ AI การพยากรณ์ด้วยวิธีการ FANN และ GA ให้ค่าวัดประสิทธิภาพดีที่สุด พิจารณาค่าวัดประสิทธิภาพ RMSE และ MAE การพยากรณ์ด้วยวิธีการ GA ให้ค่าวัดประสิทธิภาพที่ดีที่สุด ส่วนค่าวัดประสิทธิภาพ MBE วิธีการที่ให้ค่าวัดประสิทธิภาพที่ดีที่สุดคือการพยากรณ์ด้วยวิธี GA เช่นเดียวกัน

เมื่อพิจารณาโดยภาพรวมพบว่า ค่าวัดประสิทธิภาพจากทั้งสามวิธีไม่แตกต่างกันมากนัก แต่วิธีจีเนติกอัลกอริทึมให้ค่าวัดประสิทธิภาพที่ดีที่สุด พิจารณากราฟแสดงระดับค่าวัดประสิทธิภาพของการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าและวิธีจีเนติกอัลกอริทึมดังในภาพที่ 23



ภาพที่ 23 ระดับค่าวัดประสิทธิภาพของการพยากรณ์ด้วยวิธี MLR, FANN และ GA

อภิปรายผล

ในการสร้างโมเดลในการพยากรณ์อินทรีย์วัตถุที่มีประสิทธิภาพ สิ่งสำคัญที่สุดคือกระบวนการประมวลข้อมูลขั้นต้น (Data preprocessing) มีความสำคัญมากเนื่องจากการมีข้อมูลที่ไม่จำเป็นในโมเดลน้อยยิ่งเป็นผลดี ซึ่งงานวิจัยนี้ใช้วิธีวิเคราะห์องค์ประกอบหลัก ในการลดตัวแปรอิสระหรือข้อมูลที่ไม่จำเป็นออกจากโมเดล และยังคงสารสนเทศของข้อมูลไว้ ทำให้ได้ตัวแปรอิสระจำนวนที่เหมาะสมในโมเดล สามารถขจัดความสัมพันธ์ระหว่างตัวแปรอิสระ ที่อาจนำไปสู่ค่าทำนายที่ผิดพลาดได้

การพยากรณ์อินทรีย์วัตถุด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าและวิธีจีเนติกอัลกอริทึม การเลือกใช้วิธีการทั้งสามวิธีมีความแตกต่างกัน และภายใต้ข้อสมมติที่แตกต่างกัน วิธีการถดถอยเชิงเส้นพหุคูณ เป็นวิธีการที่เข้าใจง่ายและใช้งานง่าย ไม่ซับซ้อน แต่มีข้อเสียคือข้อมูลที่ใช้ต้องเป็นข้อมูลแบบเชิงเส้น เนื่องจากข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลทางสิ่งแวดล้อมที่มีการเปลี่ยนแปลงอยู่ตลอดเวลาจึงมีความไม่เป็นเชิงเส้นสูง ดังนั้นจึงเป็นข้อจำกัดในการใช้วิธีการถดถอยเชิงเส้นพหุคูณ วิธีการที่สองคือวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า เป็นวิธีการที่ไม่คำนึงถึงความเป็นเชิงเส้นของข้อมูลได้ แต่เนื่องจากวิธีการโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าจะมีความแม่นยำในการพยากรณ์มากเท่าใดนั้น ขึ้นอยู่กับกำหนดโครงข่าย จำนวนโหนดในชั้นซ่อน การกำหนดฟังก์ชันการแปลง และอัลกอริทึมในการเรียนรู้ที่เหมาะสมให้กับโมเดล เพื่อไม่ให้เกิดปัญหาตัวแบบเหมาะสมเกินไป (over fitting) และเนื่องจากการสร้างและทดสอบโมเดลไม่มีข้อกำหนดตายตัวขึ้นอยู่กับรูปแบบของปัญหา ดังนั้นจึงต้องอาศัยวิธีลองผิดลองถูก (trial and error) ส่วนวิธีสุดท้ายคือวิธีจีเนติกอัลกอริทึมซึ่งเป็นวิธีไม่ต้องคำนึงถึงความเป็นเชิงเส้นของข้อมูล และสามารถใช้กับปัญหาที่มีเงื่อนไขมากๆ หรือการจำกัดขอบเขตในการหาคำตอบ สามารถแก้ปัญหาที่ยุ่งยาก ซับซ้อนได้ ซึ่งเป็นข้อดีของการใช้วิธีจีเนติกอัลกอริทึม งานวิจัยนี้ใช้วิธีจีเนติกอัลกอริทึมร่วมกับวิธีการถดถอยเชิงเส้นพหุคูณ ในการเลือกค่าสัมประสิทธิ์ให้กับโมเดลการถดถอยเชิงเส้นพหุคูณ จึงต้องมีการกำหนดขอบเขตของตัวแปร และกำหนดพารามิเตอร์ให้เหมาะสมกับปัญหา

การพยากรณ์อินทรีย์วัตถุด้วยวิธีการถดถอยเชิงเส้นพหุคูณ วิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้าและวิธีจีเนติกอัลกอริทึม จากผลลัพธ์จะพบว่าค่าวัดประสิทธิภาพ

ของทั้งสามวิธี ให้ค่าวัดประสิทธิภาพไม่แตกต่างกันมากนัก แต่โดยภาพรวมวิธีจีเนติกอัลกอริทึมให้ผลลัพธ์ที่แม่นยำที่สุด อาจเนื่องมาจากวิธีจีเนติกอัลกอริทึมไม่ต้องคำนึงถึงความเป็นเชิงเส้นของข้อมูล แม้ว่าวิธีจีเนติกอัลกอริทึมเป็นวิธีที่เข้าใจยาก ซับซ้อน แต่ก็เป็นวิธีการใหม่ที่เหมาะสมกับข้อมูลทางด้านสิ่งแวดล้อมหรือข้อมูลที่มีการเปลี่ยนแปลงอยู่ตลอดเวลา ทำให้สามารถนำมาประยุกต์สร้างโมเดลในการพยากรณ์เพื่อให้มีประสิทธิภาพที่แม่นยำกว่าได้

ข้อจำกัดในงานวิจัย

การเปรียบเทียบประสิทธิภาพของการพยากรณ์ที่ได้จากวิธีจีเนติกอัลกอริทึมต้องมีการกำหนดพารามิเตอร์ของกระบวนการให้เหมาะสมกับขอบเขตของปัญหา เพื่อไม่ให้เกิดความแตกต่างในการเปรียบเทียบประสิทธิภาพของการพยากรณ์ที่ได้จากวิธีจีเนติกอัลกอริทึม กับวิธีถดถอยเชิงเส้นพหุคูณและวิธีโครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้า

ข้อเสนอแนะจากงานวิจัย

1. ในการเลือกอัลกอริทึมในการเรียนรู้โครงข่ายประสาทเทียมแบบที่มีการเชื่อมโยงไปข้างหน้างานวิจัยนี้ใช้ลาเวนเบิร์ก มาควอทอัลกอริทึม (Lavenberg Maquard Algorithm) ในการสอนโครงข่ายซึ่งผู้วิจัยเลือกใช้เนื่องจากอัลกอริทึมนี้เหมาะสมกับปัญหาขนาดเล็ก ในการศึกษาในอนาคตควรทดลองใช้แบ็คพรอพเพเกชั่นอัลกอริทึม (Backpropagation Algorithm) ในการสอนโครงข่าย

2. การใช้วิธีการจีเนติกอัลกอริทึมในการเลือกพารามิเตอร์ที่เหมาะสมให้กับสมการการถดถอยเชิงเส้นพหุคูณ สามารถที่จะลดจำนวนตัวแปรอิสระที่ไม่มีนัยสำคัญออกจากโมเดลก่อนที่จะเข้าสู่กระบวนการจีเนติกอัลกอริทึมได้ ซึ่งยิ่งมีตัวแปรน้อยยิ่งเป็นผลดีและทำให้โมเดลสามารถอธิบายความสัมพันธ์จริงของข้อมูลได้มากยิ่งขึ้น และในส่วนของการกำหนดพารามิเตอร์ต่างๆ ที่จำเป็นในกระบวนการจีเนติกอัลกอริทึมนั้น ไม่มีรูปแบบตายตัวขึ้นอยู่กับรูปแบบของแต่ละปัญหาจึงต้องอาศัยการปรับพารามิเตอร์ที่เหมาะสมสำหรับปัญหานั้น ๆ

บรรณานุกรม

ภาษาไทย

คณาจารย์ภาควิชาปฐพีวิทยา. ปฐพีวิทยาเบื้องต้น. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์, 2548.

กัลยา วานิชบัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. กรุงเทพฯ : สำนักพิมพ์บริษัทธรรมสารจำกัด, 2551.

ศรัลย์ ปานศรีพงษ์. “การพัฒนาแบบจำลองโครงข่ายใยประสาทประดิษฐ์และแบบจำลองความถดถอยเชิงพหุ เพื่อทำนายความเข้มข้นของ PM10 ในพื้นที่กรุงเทพมหานครและเทศบาลนครราชสีมา.” วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชา วิศวกรรมสิ่งแวดล้อม บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีสุรนารี, 2550.

อาทิยต์ ศรีแก้ว. “จีเนติกอัลกอริทึม ตอนที่ 1.” วารสารเทคโนโลยีสุรนารี 9, 1 (มกราคม-มีนาคม 2545) : 69-83.

ภาษาต่างประเทศ

Adeli, H. "Computer-aided engineering in the 1990's." The International Journal of Construction Information Technology, no.1 (1992) : 1-10.

Adineh, Vahidreza R. and others. "Optimization of the operational parameters in a fast axial flow CW CO_2 laser using artificial neural networks and genetic algorithms." Optics & Laser Technology 40, 8 (November 2008) : 1000-1007.

Artificial Neural Network [Online]. Accessed 20 November 2009. Available from http://202.28.94.55/web/320417/2548/work1/g26/Files/Report_Neural%20Network.doc

Ashraf, Elazouni M. and others. "Estimating Resource Requirements at Conceptual Design Stage Using Neural Networks." Journal of Computing in Civil Engineering 11, 4 (October 1997) : 217-223.

Bhokha, Sdhabhon. “Application of artificial neural networks to cost and duration forecasting for buildings.” Ph.D. Dissertation, Asian Institute of Technology, 1998.

Carpenter, William C., and Jean Francois Barthelemy. "Common misconceptions about neural networks as approximators." Journal of Computing in Civil Engineering 8, 3 (July 1994) : 345-358.

Chiraphadhandhanakul, Somsong, Pataya Dangprasert, and Vichit Avatchanakorn. "Genetic algorithms in forecasting commercial banks deposit." In IEEE International Conference on Intelligent Processing Systems, 116-121. Edited by IEEE. China : IEEE, 1997.

Cort, Willmott J. "Some comment on the evaluation of model performance." Bulletin American Meteorological Society 63, 11 (November 1982) : 1309-1313.

Dahe, Jiang and others. "Progress in developing an ANN model for air pollution index forecast." Atmospheric Environment 38, 40 (December 2004) : 7055-7064.

Erenturk, S., and Koksal Erenturk. "Comparison of genetic algorithm and neural network approaches for the drying process of carrot." Journal of Food Engineering 78, 3 (Febuary 2007) : 905-912.

Flood, I., and Kartam Nabil. "Neural network in civil engineering-II: System and Application." Journal of Computing in Civil Engineering 8, 2 (April 1994) : 149-162.

Haykin, S. Neural Networks: A Comprehensive Foundation. 2nd ed. New York : Prentice Hall, 1998.

Hisao-Tien, P., "A comparison of neural network and multiple regression analysis in modeling capital structure." Expert Systems with Applications 35, 3 (October 2008) : 720-727.

Ingleby, H.R., and Crowe T.G., "Reflectance models for predicting organic carbon in Saskatchewan soils." Canadian Biosystems Engineering 42, 2 (May 2000) : 57-63.

Ingleby, H.R., and Crowe T.G., "Neural network models for predicting organic matter content in Saskatchewan soils." Canadian Biosystems Engineering Vol.43 (2001) : 7.1 - 7.5

Jung Yi Kim and others. "Comparative study on artificial neural network with multiple regressions for continuous estimation of blood pressure." In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 6942-6945. Edited by IEEE. China : IEEE,1992.

Karunakar,B., and Datta G.L., "Controlling green sand mould properties using artificial neural networks and genetic algorithms - A comparison." Applied Clay Science. 37, 1-2 (June 2007) : 58-66.

Karunasekera, H.N.D. "Neural network structure generation for the classification of remotely sensed data using simulated annealing." M.Eng Thesis. Asian Institute of

Technology, 1992.

Khan, A.I., Topping B.H.V., and Bahreininejad A, "Parallel training of neural networks for finite element mesh decomposition." Neural Networks and Combinatorial Optimization in Civil and Structural Engineering 63, 4 (1997) : 693-707.

Kireetoh, S. “Neural networks technology.” In Engineering Institute of Thailand, 371-384. Edited by Engineering Institute of Thailand. Thailand : Engineering Institute of Thailand ,1995.

Klimasuaskas, C., "Applying neural networks." In Neural Networks in Finance and Investing, 47-72. Edited by Robert R. Trippi and Efraim Turban. New York : McGraw-Hill, Inc., 1993.

Kucuk I., Naim Derebasi, "Prediction of power losses in transformer cores using feed forward neural network and genetic algorithm." Measurement 39, 7 (August 2006) : 605-611.

Lippmann, Richard P. "An introduction to computing with neural nets." IEEE ASSP Magazine 16, 1 (March 1987) : 4–22.

Matignon, R. Data Mining Using SAS Enterprise Miner. 2nd ed. Hoboken, NJ : John Wiley & Sons, Inc., 2007.

McKim, R., Adas A., and Handa V.K. "Construction Firm Organizational Effectiveness : A Neural Network-based Prediction Methodology." In The Organization and Management of Construction Shaping and Practice, 247-256. Edited by David Langford and Arkady Retik. London : E&FN spon, and imprint of chapman & Hall, 247-256.

Medsker, L., Turban Efraim, and Robert R. Trippi, "Neural network fundamentals for financial analysis." The Journal of Investing 2, 1 (Spring 1993) : 59-68.

Merdun, H., Ozer Cinar, Ramazan Meral, and Mehmet Apan. “Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity.” Soil and Tillage Research 90, 1-2 (November 2006) : 108-116.

Rurkhamet, B., Parames Chutima, and Manop Reodech, "Comparative study of artificial neural network and regression analysis for forecasting new issued banknotes." Thammasat Int. Sc. Tech 3, 2 (July 1998) : 21-28.

Murray S. Neural Networks for Statistical Modeling. New York : John Wiley & Sons, Inc., 1993.

Jae, Song H. and others. "Comparative Analysis of Logistic Regression and Artiticial neural network for Computer-Aided Diagnosis of Breast Masses." Academic Radiology 12, 4 (April 2005) : 487-495.

Sousa, S.I.V. and others. "Multiple linear regression and artificial neural networks based on principal component to predict ozone concentrations." Environmental Modelling & Software 22, 1 (January 2007): 97-103.

Stevenson, F.J. Humus Chemistry. 2nd ed. New York : John Wiley & Sons, Inc., 1994.

Thomas, S., Martin Skitmore, and Keung Fai Wong. "Using genetic algorithms and linear regression analysis for private housing demand forecast." Building and Environment 43, 6 (June 2008) : 1171-1184.

Wu, X., and Lim S. Yong. "Prediction of maximum scour depth at spur dikes with adaptive neural networks." In Neural Networks and Combinatorial Optimization in Civil and Structural Engineering, 61-66. Edited by Civil-Comp Press. Edinburgh : Civil-Comp Press,1993.

Yang, L. and others. "Neural network and GA approaches for dwelling fire occurrence prediction." Knowledge-Based Systems 19, 4 (August 2006) : 213-219.

ภาคผนวก

ภาคผนวก ก
โปรแกรมการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม

โปรแกรมการพยากรณ์ด้วยวิธีจีเนติกอัลกอริทึม

ส่วนที่ 1 การกำหนดพารามิเตอร์สำหรับกระบวนการจีเนติกอัลกอริทึม

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package gamlr;

import java.io.IOException;

/**
 *
 * @author numvarn
 */
public class OM {

    static int CHROMOSOME_LENGHT;
    final static int POP_SIZE = 52;
    final static int TOURNAMENT_SIZE = 10;
    final static double CROSSOVER_PROB = 0.85;
    final static int MAX_GENERATION = 100;
    final static double minB0 = 2.376, maxB0 = 2.496;
    final static double minB1 = 0.104, maxB1 = 0.159;
    final static double minB2 = 0.319, maxB2 = 0.391;
    final static double minB3 = -0.091, maxB3 = -0.008;
    final static double minB4 = -0.037, maxB4 = 0.049;
    final static double minB5 = -0.238, maxB5 = -0.123;
    /* Real PC Data */
    static double[][] realPC;
    /* Real OM Data */
    static double[] realOM;

    /**
     * @param args[0] is Data file
     */
    public static void main(String[] args) throws IOException {
        if (args.length == 1) {
            /*Read Data from CSV File*/
            readCSV csvReader = new readCSV();
            csvReader.read(args[0]);

            /*Gent Chromosome lenght*/
            CHROMOSOME_LENGHT = csvReader.getPCFieldNumber();

            realPC = csvReader.getPCData();
            realOM = csvReader.getOMData();

            /*Run Genetic Algrithm*/
            GA ga = new GA();
            ga.evole();
        } else {
            System.out.println("Please, enter your query file.");
            System.exit(0);
        }

    }
}
```

ส่วนที่ 2 การกำหนดฟังก์ชั่นวัตถุประสงค์

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package gamlr;

/**
 *
 * @author numvarn
 */
public class fitnessFunction {

    /**
     * @param Individual
     * ind[0] is Beta-0
     * ind[1] is Beta-1
     * ind[2] is Beta-2
     * ind[3] is Beta-3
     * ind[4] is Beta-4
     * ind[5] is Beta-5
     * Om-Predict = Beta-0
     *              + (Beta-1 * PC1)
     *              + (Beta-2 * PC2)
     *              + (Beta-3 * PC3)
     *              + (Beta-4 * PC4)
     *              + (Beta-5 * PC5)
     */
    public double calculate(double[] ind) {
        double rmse = 0.0;
        double[] omPredict = new double[OM.realPC.length];

        for (int i = 0; i < OM.realPC.length; i++) {
            omPredict[i] = ind[0] //
                    + (ind[1] * OM.realPC[i][0]) //
                    + (ind[2] * OM.realPC[i][1]) //
                    + (ind[3] * OM.realPC[i][2]) //
                    + (ind[4] * OM.realPC[i][3]) //
                    + (ind[5] * OM.realPC[i][4]);
        }

        rmse = calculateMSE(omPredict);

        return rmse;
    }

    /**
     * Get RMSE
     */
    private double calculateMSE(double[] omPredict) {
        double rmse = 0.0;
        double sum = 0.0;

        for(int i = 0; i < omPredict.length; i++) {
            sum = sum + Math.pow(OM.realOM[i] - omPredict[i], 2);
        }

        rmse = Math.sqrt(sum / omPredict.length);
        //mse = sum / omPredict.length;

        return rmse;
    }
}
```

ส่วนที่ 3 กระบวนการจีเนติกอัลกอริทึม

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package gamlr;

import java.io.IOException;
import java.util.Random;
import java.util.logging.Level;
import java.util.logging.Logger;

/**
 *
 * @author numvarn
 */
public class GA {

    private Random generator = new Random();
    private fitnessFunction fitnessFunc = new fitnessFunction();
    private writeFile writer = new writeFile();
    private double[][] population = new double[OM.POP_SIZE][];
    private double[] fitness = new double[OM.POP_SIZE];
    private double[] tmp_offspring1 = new double[OM.CHROMOSOME_LENGHT];
    private double[] tmp_offspring2 = new double[OM.CHROMOSOME_LENGHT];
    private double bestFitness = 1e64;
    private int bestIndex = 0;
    private String resultText;

    /**
     * Main method for GA
     */
    public void evole() {
        int generation = 1;
        int parent_index1, parent_index2;
        inititalPopulation();

        while (generation < OM.MAX_GENERATION) {
            /*Select two parent*/
            parent_index1 = tournamentSelection();
            parent_index2 = tournamentSelection();

            if (generator.nextDouble() <= OM.CROSSOVER_PROB) {
                crossover(parent_index1, parent_index2);
                replacement();
            } else {
                tmp_offspring1 = population[parent_index1];
                tmp_offspring2 = population[parent_index2];
                replacement();
            }

            /*Record statistic before go to next generation*/
            recordStatistic(generation);

            generation++;
        }

        /**
         * Record all result to text file
         */
        try {
            writer.write(resultText);
        } catch (IOException ex) {
            Logger.getLogger(GA.class.getName()).log(Level.SEVERE, null, ex);
        }
    }

    /**
     * Initialize population
     */
```

```
private void inititalPopulation() {
    for (int i = 0; i < OM.POP_SIZE; i++) {
        /*Generate individual*/
        population[i] = generateIndividual();
        fitness[i] = fitnessFunc.calculate(population[i]);
        getBestIndividual(fitness[i], i);
    }
}

/**
 * Generte Individual
 */
private double[] generateIndividual() {
    double[] individual = new double[OM.CHROMOSOME_LENGHT];

    individual[0] = genBeta(OM.minB0, OM.maxB0);
    individual[1] = genBeta(OM.minB1, OM.maxB1);
    individual[2] = genBeta(OM.minB2, OM.maxB2);
    individual[3] = genBeta(OM.minB3, OM.maxB3);
    individual[4] = genBeta(OM.minB4, OM.maxB4);
    individual[5] = genBeta(OM.minB5, OM.maxB5);

    return individual;
}

private double genBeta(double min, double max) {
    double beta = nextDouble(min, max);
    return beta;
}

/**
 * Generate random double number in range
 */
private double nextDouble(double min, double max) {
    double result = generator.nextDouble() * (max - min) + min;
    if (result >= max) {
        result = nextDouble(min, max);
    }
    return result;
}

/**
 * Compare the best individual
 */
private void getBestIndividual(double fitness_value, int index) {
    if (bestFitness > fitness_value) {
        bestFitness = fitness_value;
        bestIndex = index;
    }
}

/**
 * Print individual
 */
private String printIndividual(double[] individual) {
    String tmp = null;
    for (int i = 0; i < individual.length; i++) {
        tmp += "Beta " + i + " : " + individual[i] + "\n";
    }

    return tmp;
}

/**
 * Selection operator
 * using Tournament method
 * @return population index
 */
private int tournamentSelection() {
```

```
    int index = 0, selectedIndex = 0;
    int min = 0, max = OM.POP_SIZE - 1;
    double minValue = 1e64;

    for (int i = 0; i < OM.TOURNAMENT_SIZE; i++) {
        index = generator.nextInt(max) + min;

        if (minValue > fitness[index]) {
            minValue = fitness[index];
            selectedIndex = index;
        }
    }

    return selectedIndex;
}

/**
 * Crossover Operation
 * result is two offsprint
 * Store result in Global variable tmp_offspring1 & tmp_offspring2
 */
private void crossover(int indexParent1, int indexParent2) {
    int crossPoint = selectCrossoverOnePoint();
    double[] parent1 = population[indexParent1];
    double[] parent2 = population[indexParent2];

    int lastLength = (OM.CHROMOSOME_LENGHT - 1) - crossPoint;

    /*Create Offsprint 1*/
    System.arraycopy(parent1, 0, tmp_offspring1, 0, crossPoint + 1);
    System.arraycopy(parent2, crossPoint + 1, tmp_offspring1, crossPoint + 1, lastLength)

    /*Create Offsprint 2*/
    System.arraycopy(parent2, 0, tmp_offspring2, 0, crossPoint + 1);
    System.arraycopy(parent1, crossPoint + 1, tmp_offspring2, crossPoint + 1, lastLength)
}

/*Select crossover point
 *using one point one point crossover
 */
private int selectCrossoverOnePoint() {
    int index = generator.nextInt(OM.CHROMOSOME_LENGHT) + 0;

    if (index == (OM.CHROMOSOME_LENGHT - 1)) {
        index = OM.CHROMOSOME_LENGHT - 2;
    } else if (index == 0) {
        index = 1;
    }

    return index;
}

/*
 * Replacement Operation
 * Using Steady State Method
 */
private void replacement() {
    int index1 = nagativeTournamentSelection();
    int index2 = nagativeTournamentSelection();

    double fitness_offspring1 = fitnessFunc.calculate(tmp_offspring1);
    double fitness_offspring2 = fitnessFunc.calculate(tmp_offspring2);

    if (fitness_offspring1 < fitness[index1]) {
        fitness[index1] = fitness_offspring1;
        population[index1] = tmp_offspring1;

        getBestIndividual(fitness[index1], index1);
    }
```

```
            if (fitness_offspring2 < fitness[index2]) {
                fitness[index2] = fitness_offspring2;
                population[index2] = tmp_offspring2;

                getBestIndividual(fitness[index2], index2);
            }
        }

        /*
         * Replacement Method
         * using Nagative Tournament
         * @return index
         */
        private int nagativeTournamentSelection() {
            int index = 0, selectedIndex = 0;
            int min = 0, max = OM.POP_SIZE - 1;
            double maxValue = 0;

            for (int i = 0; i < OM.TOURNAMENT_SIZE; i++) {
                index = generator.nextInt(max) + min;

                if (maxValue < fitness[index]) {
                    maxValue = fitness[index];
                    selectedIndex = index;
                }
            }

            return selectedIndex;
        }

        /**
         * Print Statictic
         * @param generation
         * keep static to global variable
         */
        private void recordStatistic(int generation) {
            /*Record statistic data*/
            String x = Integer.toString(generation);

            resultText = resultText + "\n\nGeneration " + x
                    + "\nBest RMSE is :\t" + fitness[bestIndex] + "\n" + bestIndex + "\n "
                    + printIndividual(population[bestIndex]);
        }
    }
```

ส่วนที่ 4 การนำเข้าไฟล์

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package gamlr;

import java.io.DataInputStream;
import java.io.FileInputStream;
import java.io.IOException;
import java.util.StringTokenizer;

public class readCSV { //readCSV class starts here

    private int countLine = 0;
    private int countField = 0;
    private String fileName;

    public void read(String filePath) throws IOException {
        fileName = filePath;
        //string variable which take each record at a time
        String thisLine;

        //A FileInputStream obtains input bytes from a file in a file system
        FileInputStream fis = new FileInputStream(fileName);

        /*om input stream lets an application read primitive Java om types
        from an underlying input stream in a machine-independent way*/
        DataInputStream myInput = new DataInputStream(fis);

        while ((thisLine = myInput.readLine()) != null) {
            //beginning of outer while loop
            StringTokenizer st = new StringTokenizer(thisLine, ",");
            countField = 0;
            while (st.hasMoreElements()) {
                String field = st.nextToken();
                //System.out.print(field + ", ");
                countField++;
            }
            countLine++;
        } //ending of outer while loop
    }

    public double[][] getPCData() throws IOException {
        double[][] dataPC = new double[countLine][countField - 1];
        String thisLine;

        FileInputStream fis = new FileInputStream(fileName);
        DataInputStream myInput = new DataInputStream(fis);

        int line_no = 0;
        int field_no = 0;
        while ((thisLine = myInput.readLine()) != null) {
            StringTokenizer st = new StringTokenizer(thisLine, ",");
            field_no = 0;
            while (st.hasMoreElements()) {
                String field = st.nextToken();
                if (field_no < countField - 1) {
                    dataPC[line_no][field_no] = Double.parseDouble(field);
                }
                field_no++;
            }
            line_no++;
        }
        return dataPC;
    }

    public double[] getOMData() throws IOException {
        double[] om = new double[countLine];
        String thisLine;
```

```
        FileInputStream fis = new FileInputStream(fileName);
        DataInputStream myInput = new DataInputStream(fis);

        int line_no = 0;
        int field_no = 0;
        while ((thisLine = myInput.readLine()) != null) {
            StringTokenizer st = new StringTokenizer(thisLine, ",");
            field_no = 0;
            while (st.hasMoreElements()) {
                String field = st.nextToken();
                if (field_no == countField - 1) {
                    om[line_no] = Double.parseDouble(field);
                }
                field_no++;
            }
            line_no++;
        }
        return om;
    }

    /**
     * Return number of PC Field
     */
    public int getPCFieldNumber() {
        return countField;
    }

} //readCSV class ends here
```

ส่วนที่ 5 การอ่านผลลัพธ์

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package gamlr;

import java.io.BufferedWriter;
import java.io.File;
import java.io.FileWriter;
import java.io.IOException;
import java.io.Writer;
import java.util.logging.Level;
import java.util.logging.Logger;

/**
 *
 * @author numvarn
 */
public class writeFile {

    File file;
    Writer output = null;

    public writeFile() {
        try {
            file = new File("result.txt");
            output = new BufferedWriter(new FileWriter(file));
        } catch (IOException ex) {
            Logger.getLogger(writeFile.class.getName()).log(Level.SEVERE, null, ex);
        }
    }

    public void write(String strInput) throws IOException {
        output.write(strInput);
        output.close();
    }
}
```

ประวัติผู้วิจัย

ชื่อ – สกุล นางสาวสิริกัลยา ประมวล

ที่อยู่ 69 หมู่ที่ 4 ตำบลสมอ อำเภอปรางค์กู่ จังหวัดศรีสะเกษ

ประวัติการศึกษา

พ.ศ. 2548 สำเร็จการศึกษาระดับปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาสถิติ มหาวิทยาลัยศิลปากร

พ.ศ. 2550 ศึกษาต่อระดับปริญญามหาบัณฑิต สาขาวิชาสถิติประยุกต์ มหาวิทยาลัยศิลปากร