

การตรวจจับความสัมพันธ์เชิงพันธุกรรมโดยข่ายงานระบบประสาทเชิงวิวัฒนาการ

นางสาวพรเพ็ญ อาชีวะเกษะ

นางสาวอรุณรัตน์ อมะถัยเจียร

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปีการศึกษา 2557

Detection of Genetic Association by an Evolutionary Neural Network


Ms. Pornphen Ahcheewakesa

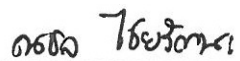
Ms. Aroonrat Amalashthira

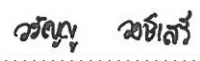
A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF BACHELOR OF COMPUTER ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S UNIVERSITY OF TECHNOLOGY NORTH BANGKOK
ACADEMIC YEAR 2014

ปริญญานิพนธ์เรื่อง : การตรวจจับความสัมพันธ์เชิงพันธุกรรมโดยข่ายงานระบบ
ประสาทเชิงวิวัฒนาการ
ชื่อ : นางสาวพรเพ็ญ อาชีวะเกษะ
นางสาวอรุณรัตน์ อมะลย์เจียร
สาขาวิชา : วิศวกรรมคอมพิวเตอร์
ภาควิชา : วิศวกรรมไฟฟ้าและคอมพิวเตอร์
คณะ : วิศวกรรมศาสตร์
อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.ณชด ไชยรัตน์
ปีการศึกษา : 2557

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ อนุมัติให้
ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์


..... หัวหน้าภาควิชาวิศวกรรมไฟฟ้า
(ผู้ช่วยศาสตราจารย์ ดร. ณชด วิวัชรโกเศศ) และคอมพิวเตอร์


..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.ณชด ไชยรัตน์)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. วรรณ วังษ์เสรี)


..... กรรมการ
(รองศาสตราจารย์ ดร. วรา วราวิทย์)

ลิขสิทธิ์ของภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Projected Report Title : Detection of Genetic Association by an Evolutionary Neural Network

Name : Ms. Pornphen Ahcheewakesa
Ms. Aroonrat Amalashthira

Major Field : Computer Engineering

Department : Electrical and Computer Engineering

Faculty : Engineering

Project Advisor : Assoc. Prof. Dr. Nachol Chaiyaratana

Academic Year : 2014

Accepted by the Faculty of Engineering, King Mongkut's University of Technology North Bangkok in Partial Fulfillment of the Requirement for the Degree of Bachelor of Computer Engineering

.....
(Asst. Prof. Dr. Nophadon Wiwatcharagoses) Chairperson of Department of Electrical and Computer Engineering

Nachol Chaiyaratana
.....
(Assoc. Prof. Dr. Nachol Chaiyaratana) Chairperson

Waranyu Wongseree
.....
(Asst. Prof. Dr. Waranyu Wongseree) Member

P.V.
.....
(Assoc. Prof. Dr. Vara Varavithya) Member

Copyright of the Department of Electrical and Computer Engineering, Faculty of Engineering
King Mongkut's University of Technology North Bangkok

บทคัดย่อภาษาไทย

ปริญญานิพนธ์ฉบับนี้นำเสนอการประยุกต์ใช้ข่ายงานระบบประสาทเชิงวิวัฒนาการสำหรับการศึกษาความสัมพันธ์เชิงพันธุกรรม โดยชนิดข่ายงานระบบประสาทที่เลือกใช้คือ มัลติเลเยอร์เพอร์เซ็ปตรอน ที่รับสนิปเป็นข้อมูลเข้าและให้สถานะกรณี – ความคุมเป็นข้อมูลออก สถาปัตยกรรมของข่ายงานซึ่งประกอบด้วยจำนวนข้อมูลเข้า, จำนวนของชั้นซ่อน และจำนวนปมซ่อนของแต่ละชั้นซ่อน ได้รับการปรับโดยใช้ขั้นตอนวิธีทางพันธุกรรมเพื่อเพิ่มประสิทธิภาพการจำแนกให้มีความแม่นยำสูง ดังนั้นข่ายงานระบบประสาทเชิงวิวัฒนาการจึงสามารถระบุสนิปตัวที่เกี่ยวข้องในการส่งผลให้เกิดโรค ข่ายงานระบบประสาทเชิงวิวัฒนาการได้รับการทดสอบด้วยปัญหามาตรฐานทั้งหมด 12 ปัญหา ในแต่ละปัญหาประกอบด้วย 10 สนิปโดยกำหนดให้มี 2 สนิปที่ส่งผลให้เกิดโรค ผลการทดลองพบว่าข่ายงานระบบประสาทเชิงวิวัฒนาการมีความสามารถในการระบุสนิปที่ส่งผลให้เกิดโรค โดยงานในอนาคตเพื่อที่จะพัฒนาประสิทธิภาพของข่ายงานระบบประสาทเชิงวิวัฒนาการจะถูกยกมาพูดถึงด้วย

บทคัดย่อภาษาอังกฤษ

This thesis presents an application of evolutionary neural network for genetic association studies. The neural network of interest is a multilayer perceptron that takes single nucleotide polymorphisms (SNPs) as inputs and gives case-control status as an output. The network architecture including the number of inputs, the number of hidden layers, and the number of hidden nodes in each hidden layer is optimized to achieve the highest classification accuracy using a genetic algorithm. This provides a means for identifying SNPs that are associated with the disease. The evolutionary neural network was tested on 12 benchmark problems involving 10 SNPs where 2 SNPs are causative. The results indicate that the evolutionary neural network was capable of identifying causative SNPs in most problems. Further works to improve the performance of evolutionary neural network were also discussed.

กิตติกรรมประกาศ

ปริญญานิพนธ์เล่มนี้ข้าพเจ้าได้ทุ่มเทความสามารถเพื่อให้ผลงานนี้มีคุณภาพและมีประโยชน์สูงสุด ปริญญานิพนธ์นี้ไม่อาจสำเร็จได้โดยปราศจากบุคคลที่คอยให้ความช่วยเหลือ ต้องขอขอบพระคุณบุคคลเหล่านั้นมา ณ โอกาสนี้

กราบขอบพระคุณ รองศาสตราจารย์ ดร.ณชล ไชยรัตน์ อาจารย์ที่ปรึกษาปริญญานิพนธ์ที่ได้ให้ความรู้ คำปรึกษา คำแนะนำ ข้อคิดเห็น ข้อมูล และการสนับสนุนอย่างเต็มที่ ซึ่งเป็นประโยชน์อย่างยิ่งสำหรับปริญญานิพนธ์เล่มนี้

กราบขอบพระคุณอาจารย์ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ทุกท่านที่ได้ให้วิชาความรู้ตลอด 4 ปีที่ผ่านมา ทำให้ข้าพเจ้าสามารถนำความรู้ที่เรียนมาประยุกต์ใช้ในปริญญานิพนธ์ได้เป็นอย่างดี

กราบขอบพระคุณคุณพ่อและคุณแม่ที่ให้การสนับสนุนและความรักความห่วงใยข้าพเจ้าตลอดมา ทำให้ข้าพเจ้าประสบความสำเร็จไปด้วยดี

ขอบคุณรุ่นพี่ รุ่นน้อง และเพื่อนทุกคนสำหรับคำแนะนำและความช่วยเหลือในทุกเรื่องทำให้ข้าพเจ้าทำงานได้อย่างมีประสิทธิภาพ

พรเพ็ญ อาชีวะเกษะ

อรุณรัตน์ อมะลย์เจียร

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	จ
บทคัดย่อภาษาอังกฤษ	ฉ
กิตติกรรมประกาศ	ช
สารบัญตาราง	ญ
สารบัญภาพ	ฎ
สารบัญภาพ (ต่อ)	ฏ
บทที่ 1. บทนำ	1
บทที่ 2. วิธีการและเทคนิคที่ใช้ในการทดลอง	2
2.1 ข่ายงานระบบประสาท (Neural Network)	2
2.2 ข่ายงานแบบป้อนไปข้างหน้าหลายชั้น (Multilayer Feed-forward Network)	4
2.3 บทสรุปของขั้นตอนวิธีแพร่ย้อนหลัง (Summary of the Back-Propagation Algorithm)	5
2.4 ขั้นตอนวิธีเชิงพันธุกรรมอย่างง่าย (Simple Genetic Algorithm)	7
2.5 10-fold Cross Validation	12
บทที่ 3. การทำงานและรายละเอียดของชุดข้อมูลที่ใช้ในการทดลอง	13
3.1 การทำงานของโปรแกรม	13
3.2 ข้อมูลที่ใช้ในการทดลอง	14
บทที่ 4. ผลการทดลองและการวิเคราะห์ผลการทดลอง	21
4.1 การทดลอง	21
4.2 ผลการทดลอง	21
4.3 ตารางเวลาในการทดลอง	46
บทที่ 5. สรุปผลการวิจัย	47
เอกสารอ้างอิง	48

สารบัญ (ต่อ)

	หน้า
ประวัติผู้แต่ง	50

สารบัญตาราง

ตารางที่	หน้า
3-1 Penetrance ของปัญหา Epistasis 1	15
3-2 Penetrance ของปัญหา Epistasis 2	15
3-3 Penetrance ของปัญหา Epistasis 3	16
3-4 Penetrance ของปัญหา Epistasis 4	16
3-5 Penetrance ของปัญหา Epistasis 5	16
3-6 Penetrance ของปัญหา Epistasis 6	17
3-7 Penetrance ของปัญหา Heterogeneity 1	17
3-8 Penetrance ของปัญหา Heterogeneity 2	17
3-9 Penetrance ของปัญหา Heterogeneity 3	18
3-10 Penetrance ของปัญหา S1	18
3-11 Penetrance ของปัญหา S2	18
3-12 Penetrance ของปัญหา S3	19
4-1 เวลาในการทดลองของทั้ง 12 ปัญหา	46

สารบัญภาพ

ภาพที่	หน้า
2-1 แผนภาพของมัลติเลเยอร์เพอร์เซ็ปตรอน	2
2-2 กราฟข้อมูลออกของเพอร์เซ็ปตรอนเมื่อใช้ฟังก์ชันเชิงตรรกะเป็นฟังก์ชันการกระตุ้น	3
2-3 แผนภาพแสดงการเปลี่ยนสถานะของสัญญาณภายในนิวรอน	3
2-4 ข่ายงานแบบป้อนไปข้างหน้าหลายชั้นที่มีชั้นซ่อนหนึ่งชั้นเป็นข่ายงานแบบเชื่อมโยงเต็มอัตรา (Fully connected network)	4
2-5 สายรหัสแทนต้นทางข้อมูลเข้าและลักษณะประจำ จำนวนชั้นซ่อน จำนวนนิวรอนในชั้นซ่อนและตัวถ่วงน้ำหนักการเชื่อมโยงของข่ายงานระบบประสาท	8
2-6 การคัดเลือกโดยการชักตัวอย่างทุกตัวแบบเป็นสุ่ม (Stochastic universal sampling selection)	10
2-7 รหัสเทียมสำหรับการคำนวณค่าสัมประสิทธิ์ที่ใช้ในการปรับมาตราความแข็งแรง	11
3-1 รหัสเทียมสำหรับการคำนวณค่าสัมประสิทธิ์ที่ใช้ในการปรับมาตราความแข็งแรง	14
4-13 ตัวอย่างข้อมูลในตอนแรก	19
4-14 ตัวอย่างข้อมูลหลังจากทำการเรียงใหม่	20
4-15 ตัวอย่างแฟ้มข้อมูลตาราง Penetrance	20
4-16 ผลการทดลองจากปัญหา Epistasis 1	22
4-17 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 1	23
4-18 ผลการทดลองจากปัญหา Epistasis 2	24
4-19 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 2	25
4-20 ผลการทดลองจากปัญหา Epistasis 3	26
4-21 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 3	27
4-22 ผลการทดลองจากปัญหา Epistasis 4	28
4-23 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 4	29

สารบัญรูปภาพ (ต่อ)

ภาพที่	หน้า
4-24 ผลการทดลองจากปัญหา Epistasis 5	30
4-25 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 5	31
4-26 ผลการทดลองจากปัญหา Epistasis 6	32
4-27 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 6	33
4-28 ผลการทดลองจากปัญหา Heterogeneity 1	34
4-29 กราฟแสดงอัตราความสำเร็จของปัญหา Heterogeneity 1	35
4-30 ผลการทดลองจากปัญหา Heterogeneity 2	36
4-31 กราฟแสดงอัตราความสำเร็จของปัญหา Heterogeneity 2	37
4-32 ผลการทดลองจากปัญหา Heterogeneity 3	38
4-33 กราฟแสดงอัตราความสำเร็จของปัญหา Heterogeneity 3	39
4-34 ผลการทดลองจากปัญหา S1	40
4-35 กราฟแสดงอัตราความสำเร็จของปัญหา S1	41
4-36 ผลการทดลองจากปัญหา S2	42
4-37 กราฟแสดงอัตราความสำเร็จของปัญหา S2	43
4-38 ผลการทดลองจากปัญหา S3	44
4-39 กราฟแสดงอัตราความสำเร็จของปัญหา S3	45

บทที่ 1

บทนำ

ข่ายงานระบบประสาทเชิงวิวัฒนาการ (Evolutionary Neural Network) คือ แบบจำลองที่ประกอบด้วยข่ายงานระบบประสาท (Neural Network) และขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) โดยข่ายงานระบบประสาทจะใช้ในการสร้างแบบจำลองและใช้ขั้นตอนวิธีเชิงพันธุกรรมในการกำหนดโครงสร้างของแบบจำลอง นอกจากนี้ยังใช้เทคนิค 10-fold cross validation เพื่อเพิ่มความแม่นยำของแบบจำลอง โดยเป้าหมายของโครงงานคือการใช้ขั้นตอนวิธีเชิงพันธุกรรมในการกำหนดโครงสร้างของข่ายงานระบบประสาทอย่างเช่น attribute, จำนวน hidden layer และ จำนวน hidden node ในแต่ละ hidden layer แล้วนำข่ายงานระบบประสาทนี้ไปทำการฝึกสอนและแบ่งประเภทของผลลัพธ์ที่ได้จากข่ายงานระบบประสาทแล้วจึงนำไปหาความแม่นยำเพื่อใช้ในการระบุ Single Nucleotide Polymorphism (SNP) หรือสปีดที่ส่งผลให้เกิดโรค

ข่ายงานระบบประสาทเชิงวิวัฒนาการใช้ในการตรวจหา SNP ที่ส่งผลให้เกิดโรค โดยการทดสอบความแม่นยำของแบบจำลองจะใช้ข้อมูลจากโปรแกรมจีโนมซิมในการสร้างชุดข้อมูลทั้งหมด 12 ปัญหาตามตารางเพนแทรนซ์ โดยในแต่ละปัญหาจะมี SNP ที่ส่งผลให้เกิดโรคอยู่ 2 ตำแหน่ง

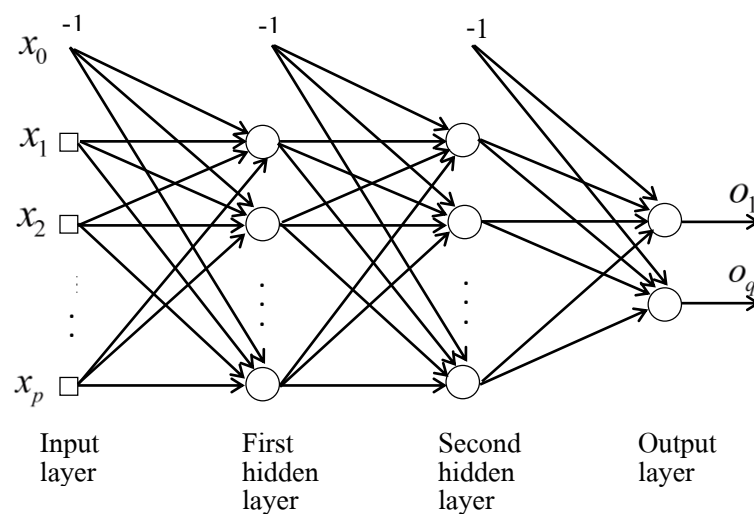
เนื้อหาของปริญญานิพนธ์ประกอบด้วย 4 ส่วนด้วยกัน ส่วนแรกคือบทที่ 2 จะกล่าวถึงวิธีการและเทคนิคที่ใช้ในการทดลอง ส่วนบทที่ 3 เป็นการทำงานและรายละเอียดของชุดข้อมูลที่ใช้ในการทดลอง บทที่ 4 เป็นผลการทดลองและการวิเคราะห์ผลการทดลอง และบทที่ 5 เป็นส่วนของการสรุปผลการทดลองและนำเสนอแนวทางในการพัฒนางาน

บทที่ 2

วิธีการและเทคนิคที่ใช้ในการทดลอง

2.1 ข่ายงานระบบประสาท (Neural Network)

ข่ายงานระบบประสาทเป็นเครื่องมือในการสร้างแบบจำลองทางคณิตศาสตร์ (Mathematical Modelling Tool) ซึ่งได้อิงแบบโครงสร้างมาจากระบบประสาท (Nervous System) (Aleksander and Morton, 1990; Haykin, 1994) โดยโครงสร้างของแบบจำลองที่เลือกใช้เป็น โครงสร้างของมัลติเลเยอร์เพอร์เซ็ปตรอน (Structure of a Multilayer Perceptron) แสดงไว้ในภาพที่ 2-1



ภาพที่ 2-1 แผนภาพของมัลติเลเยอร์เพอร์เซ็ปตรอน

โดยนิวรอนจะประกอบด้วย 2 ส่วน คือ ตัวถ่วงน้ำหนักการเชื่อมโยง (Connection Weight) และ ฟังก์ชันการกระตุ้น (Activation Function)

2.1.1 ตัวถ่วงน้ำหนักการเชื่อมโยง หรือ ตัวถ่วงน้ำหนักไซแนปส์ (Synaptic Weight) เป็นส่วนที่ได้มาเมื่อการเรียนรู้ได้เสร็จสิ้นลง โดยที่ขั้นตอนวิธีการเรียนรู้ (Learning Algorithm) จะมีหน้าที่

ทำให้ตัวถ่วงน้ำหนักการเชื่อมโยงมีการเปลี่ยนแปลงจากสถานะเริ่มต้น (Initial State) ไปเป็นสถานะสุดท้าย (Final State)

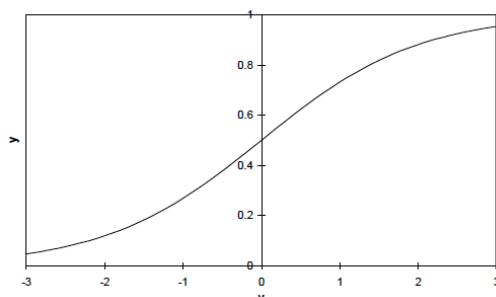
2.1.2 ฟังก์ชันการกระตุ้น คือ ฟังก์ชันภายในนิวรอนซึ่งมีหน้าที่เปลี่ยนระดับการกระตุ้นภายในสุทธิเป็นข้อมูลออกของนิวรอน (Neuron Output) โดยความสัมพันธ์ระหว่างระดับการกระตุ้นภายในสุทธิและข้อมูลออกของนิวรอนจะมีลักษณะเป็น

$$y = \phi(v) \quad (2-1)$$

โดยที่ y คือ ข้อมูลออกของนิวรอน v คือ ระดับการกระตุ้นภายในสุทธิ และ $\phi(v)$ คือ ฟังก์ชันการกระตุ้น

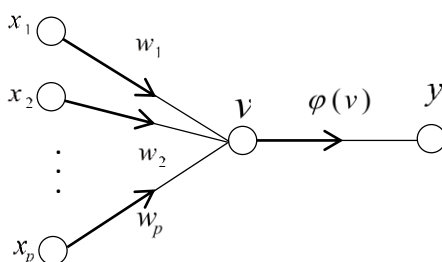
สำหรับกรณีที่ฟังก์ชันการกระตุ้นเป็นฟังก์ชันเชิงตรรกะ (Logistic Function) ตามที่แสดงในภาพที่ 2-2 ข้อมูลออกของเพอร์เซปตรอนก็จะมีค่าเป็น

$$y = \phi(v) = \frac{1}{1 + \exp(-v)}, \quad -\infty < v < \infty \quad (2-2)$$



ภาพที่ 2-2 กราฟข้อมูลออกของเพอร์เซปตรอนเมื่อใช้ฟังก์ชันเชิงตรรกะเป็นฟังก์ชันการกระตุ้น

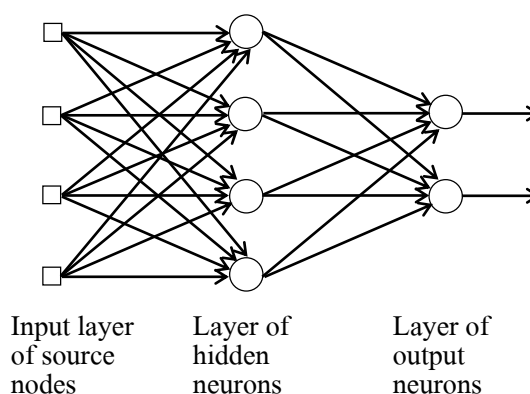
กระบวนการที่นำไปสู่การเปลี่ยนจากข้อมูลเข้าของนิวรอนไปเป็นข้อมูลออกของนิวรอน ซึ่งเกิดขึ้นภายในนิวรอนได้แสดงไว้ในภาพที่ 2-3



ภาพที่ 2-3 แผนภาพแสดงการเปลี่ยนสถานะของสัญญาณภายในนิวรอน

2.2 ข่ายงานแบบป้อนไปข้างหน้าหลายชั้น (Multilayer Feed-forward Network)

ข่ายงานแบบป้อนไปข้างหน้าหลายชั้นเป็นระบบที่มีหน้าที่ในการนำข้อมูลออกจากชั้นก่อนหน้ามาเป็นข้อมูลเข้าของนิวรอนในชั้นถัดไปตามแผนภาพที่แสดงไว้ในภาพที่ 2-4



ภาพที่ 2-4 ข่ายงานแบบป้อนไปข้างหน้าหลายชั้นที่มีชั้นซ่อนหนึ่งชั้นเป็นข่ายงานแบบเชื่อมโยงเต็มอัตรา (Fully connected network)

การเรียนรู้โดยการแก้ข้อผิดพลาด (Error-Correction Learning) คือ การเปลี่ยนแปลงค่าตัวถ่วงน้ำหนักการเชื่อมโยงภายในข่ายงานระบบประสาท (Kawato et al., 1988)

$$e_k(n) = d_k(n) - y_k(n) \quad (2-3)$$

โดยที่ $d_k(n)$ คือ ค่าข้อมูลที่ต้องการให้ออกจากนิวรอนที่ k ณ ขั้นตอนเวลา (Time Step) n $y_k(n)$ คือ ค่าข้อมูลจริงที่ออกจากนิวรอนสัญญาณข้อผิดพลาด และ $e_k(n)$ คือ ค่าความผิดพลาด ณ ขั้นตอนเวลา n

$$\varepsilon(n) = \frac{1}{2} \sum_k e_k^2(n) \quad (2-4)$$

โดยที่ $\varepsilon_k(n)$ คือ ฟังก์ชันค่าใช้จ่ายที่ขั้นตอนเวลา n อาศัยการหาค่าต่ำสุด (Minimisation) ของฟังก์ชันค่าใช้จ่ายโดยเทียบจากตัวถ่วงน้ำหนักการเชื่อมโยง

วิธีลดตามความชัน (Gradient Descent Method) และการปรับค่าตัวถ่วงน้ำหนักจะสามารถทำได้จาก

$$w(n+1) = w(n) - \eta \frac{\partial \varepsilon(n)}{\partial w} \quad (2-5)$$

โดยที่ $w(n)$ คือ ตัวถ่วงน้ำหนักการเชื่อมโยงที่ขึ้นเวลา n $\eta \frac{\partial \varepsilon(n)}{\partial w}$ คือ การหาค่าต่ำสุด (Minimization) ของฟังก์ชันค่าใช้จ่ายโดยเทียบกับตัวถ่วงน้ำหนักการเชื่อมโยง และ n คือ พารามิเตอร์อัตราการเรียนรู้ (Learning Rate Parameter)

2.3 บทสรุปของขั้นตอนวิธีแพร่ย้อนหลัง (Summary of the Back-Propagation Algorithm)

ในการปรับตัวถ่วงน้ำหนักการเชื่อมโยงภายในข่ายงาน โดยใช้ขั้นตอนวิธีแพร่ย้อนหลัง (Rumelhart and McClelland, 1986) จะสามารถอธิบายได้ดังนี้

2.3.1 Initialisation คือ การเลือกโครงสร้างของมัลติเลเยอร์เพอร์เซ็ปตรอน ซึ่งประกอบด้วย จำนวนของปมข้อมูลเข้า (Input Node) จำนวนของชั้นซ่อน (Hidden Layer) จำนวนของปมซ่อน (Hidden Node) ในแต่ละชั้นซ่อน และจำนวนของปมข้อมูลออก (Output Node) โดยที่ตัวถ่วงน้ำหนักการเชื่อมโยงและค่าขีดเริ่มเปลี่ยนในข่ายงานก็จะถูกกำหนดค่าขึ้นอย่างสุ่ม

2.3.2 Presentation of Training Examples คือ การป้อนข้อมูลที่จะใช้ในการฝึกสอนเข้าไปในข่ายงาน

2.3.3 Forward Computation คือ ขั้นตอนในการคำนวณค่าข้อมูลจริงที่ออกจากระบบเพื่อนำไปหาความผิดพลาดหรือผลต่างระหว่างค่าข้อมูลจริงกับค่าข้อมูลที่ต้องการ โดยเริ่มจาก

$$v_j^{(l)}(n) = \sum_{i=0}^p w_{ji}^{(l)}(n) y_i^{(l-1)}(n) \quad (2-6)$$

โดยที่ $y_i^{(l-1)}$ คือ ข้อมูลออกของนิวรอนที่ i ในชั้นก่อน (Previous Layer) ที่ $l-1$ ที่การทำซ้ำครั้งที่ n $w_{ji}^{(l)}(n)$ คือ ตัวถ่วงน้ำหนักระหว่างนิวรอนที่ j ในชั้นที่ l และนิวรอนที่ i ในชั้นที่ $l-1$ $v_j^{(l)}(n)$ คือ ระดับการกระตุ้นภายในสุทธิของนิวรอนที่ j และ p คือ จำนวนของนิวรอนในชั้นที่ $l-1$

ในกรณีที่ $i = 0$

$y_0^{(l-1)}$ จะมีค่าเท่ากับ -1 และ $w_{j0}^{(l)}(n)$ จะมีค่าเท่ากับ $\theta_j^{(l)}(n)$ ซึ่งก็คือค่าขีดเริ่มเปลี่ยนของนิวรอนที่ j ในชั้นที่ l โดยที่ฟังก์ชันเชิงตรรกะเป็นฟังก์ชันการกระตุ้นในแต่ละนิวรอน ข้อมูลออกจะมีค่าเป็น

$$y_j^{(l)}(n) = \frac{1}{1 + \exp(-v_j^{(l)}(n))} \quad (2-7)$$

โดยที่ $y_j^{(l)}(n)$ คือ ข้อมูลออกของนิวรอนที่ j ในชั้นที่ l ที่การทำซ้ำครั้งที่ n

ถ้านิวรอนที่ j อยู่ในชั้นซ่อนที่หนึ่ง $l=1$

$$y_i^{(0)}(n) = x_i(n), \quad i = 1, 2, \dots, p \quad (2-8)$$

โดยที่ $x_i(n)$ คือ สมาชิกที่ i ของเวกเตอร์ข้อมูลเข้า $x(n)$ p คือ จำนวนของปมข้อมูลเข้าในข่ายงาน และ $y_i^{(0)}(n)$ คือ ข้อมูลเข้าสู่นิวรอนที่ j จากปมข้อมูลเข้าที่ i

ถ้านิวรอนที่ j อยู่ในชั้นข้อมูลออก $l = L$

$$y_j^{(L)}(n) = o_j(n), \quad i = 1, 2, \dots, q \quad (2-9)$$

โดยที่ $o_j(n)$ คือ ข้อมูลจริงที่ออกจากปมข้อมูลออกที่ j q คือ จำนวนของปมข้อมูลออก และ $y_j^{(L)}(n)$ คือ ข้อมูลออกของนิวรอนที่ j ในชั้นข้อมูลออก L

ค่าของสัญญาณข้อผิดพลาดจะสามารถคำนวณได้จาก

$$e_j(n) = d_j(n) - o_j(n), \quad j = 1, 2, \dots, q \quad (2-10)$$

โดยที่ $d_j(n)$ คือ ค่าข้อมูลที่ต้องการให้ออกจากนิวรอนที่ j ณ ขั้นตอนเวลา (Time Step) n และ $e_j(n)$ คือ ค่าความผิดพลาดจากนิวรอนข้อมูลออกที่ j

2.3.4 Backward Computation คือ ขั้นตอนการปรับค่าตัวถ่วงน้ำหนักการเชื่อมโยง โดยใช้ค่าความผิดพลาดในการคำนวณ

ถ้านิวรอนที่ j อยู่ในชั้นข้อมูลออก ค่าความชันเฉพาะที่ของนิวรอนที่ j จะมีค่าเป็น

$$\delta_j^{(L)}(n) = e_j(n) o_j(n) [1 - o_j(n)] \quad (2-11)$$

โดยที่ $\delta_j^{(L)}(n)$ คือ ค่าความชันเฉพาะที่ของนิวรอนที่ j ในชั้นข้อมูลออก L

ถ้านิวรอนที่ j อยู่ในชั้นซ่อนที่ l ค่าความชันเฉพาะที่ของนิวรอนที่ j จะมีค่าเป็น

$$\delta_j^{(l)}(n) = y_j^{(l)}(n) [1 - y_j^{(l)}(n)] \sum_k \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n) \quad (2-12)$$

โดยที่ $\delta_k^{(l+1)}(n)$ คือ ค่าความชันเฉพาะที่ของนิวรอนที่ k ในชั้น $l+1$ และ $w_{kj}^{(l+1)}(n)$ คือ ตัวถ่วงน้ำหนักการเชื่อมโยงระหว่างนิวรอนที่ k ในชั้น $l+1$ กับนิวรอนที่ j ในชั้น l

การปรับตัวถ่วงน้ำหนักการเชื่อมโยงโดยใช้การเรียนรู้แบบแก้ข้อผิดพลาดจะสามารถทำได้โดย

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha [w_{ji}^{(l)}(n) - w_{ji}^{(l)}(n-1)] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n) \quad (2-13)$$

โดยที่ η คือ พารามิเตอร์อัตราการเรียนรู้ (Learning Rate Parameter) และ α คือ ค่าคงตัวโมเมนตัม (Momentum Constant)

2.3.5 Iteration คือ การคำนวณซ้ำในขั้นตอนที่ 2, 3 และ 4 เป็นจำนวน m รอบเวลา จนกระทั่งค่าตัวถ่วงน้ำหนักการเชื่อมโยงลู่เข้า สังเกตได้จากการลู่เข้าของค่าเฉลี่ยของผลรวมค่าผิดพลาดกำลังสอง (Mean of Sum of Squared Errors) จากปมข้อมูลออก ซึ่งคำนวณโดยใช้ค่าผิดพลาดระหว่างการเรียนรู้ใน 1 รอบเวลา สามารถคำนวณได้โดย

$$\text{Mean of Sum of Square Errors} = \frac{\sum_{j=1}^N \sum_{i=1}^q (e_i(j))^2}{N} \quad (2-14)$$

โดยที่ $e_i(j)$ คือ สัญญาณค่าผิดพลาดจากปมข้อมูลออกที่ i ในการทำซ้ำครั้งที่ j ของรอบเวลา

2.4 ขั้นตอนวิธีเชิงพันธุกรรมอย่างง่าย (Simple Genetic Algorithm)

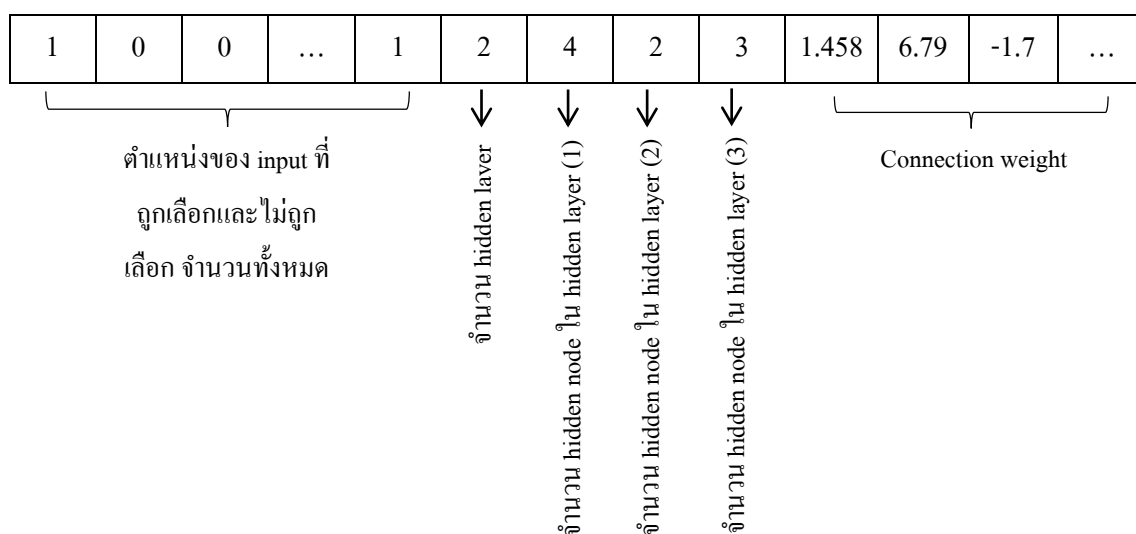
ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) เป็นการหาคำตอบที่ดีที่สุด ใน 1 รุ่น (Generation) โดยเป็นการหาคำตอบแบบสุ่มสุ่ม (Stochastic Search) หรือเทคนิคการหาค่าเหมาะที่สุด (Optimisation Technique) เพื่อทำให้เกิดคำตอบที่ดีขึ้นในรุ่นถัดไป ซึ่งกระบวนการหาคำตอบนั้นอิงมาจากแนวความคิดของชาร์ลส์ ดาร์วิน (Charles Darwin) ในเรื่องของการอยู่รอดของผู้ที่แข็งแรงที่สุด (Survival of the Fittest) (Holland, 1975; Goldberg, 1989)

ขั้นตอนวิธีเชิงพันธุกรรมอย่างง่าย (Goldberg, 1989) สามารถอธิบายได้ดังนี้

1. สร้างประชากรเริ่มต้น (Initial Population) โดยใช้โครโมโซมที่ประกอบขึ้นจากอักขรที่เลือกขึ้นอย่างสุ่ม
2. ถอดรหัสโครโมโซมของแต่ละสมาชิกในประชากรเพื่อให้ได้ตัวแปรตัดสินใจ
3. คำนวณค่าจุดประสงค์ (Objective Value) ของแต่ละผลเฉลยโดยใช้ตัวแปรตัดสินใจที่ได้เป็นข้อมูลเข้า (Input) ของฟังก์ชันจุดประสงค์
4. หาค่าความแข็งแรง (Fitness Value) ของสมาชิกแต่ละตัวโดยใช้ค่าจุดประสงค์
5. สร้างประชากรใหม่จากประชากรปัจจุบัน โดยใช้ค่าความแข็งแรงในการกำหนดว่าควรเลือกสมาชิกตัวไหนจากประชากรปัจจุบันมาไว้ในประชากรใหม่
6. ทำการแปลงสมาชิกในประชากรใหม่โดยการครอสโอเวอร์ (Crossover) และการกลายพันธุ์ (Mutation)
7. กลับไปที่ข้อ 2 จนกระทั่งผลเฉลยที่ได้มีลักษณะลู่เข้า (Converge) หรือได้ทำการคำนวณจนถึงจำนวนรุ่นสูงสุดที่กำหนดไว้

2.4.1 สร้างประชากรเริ่มต้น (Initial Population)

โดยใช้โครโมโซมที่ประกอบขึ้นเป็นตัวแทนของ Neural network ซึ่งสามารถอธิบายได้ตามภาพที่ 2-5 ตัวอย่างของ population ในรูปของ string คือ 1,0,0,1,1,1,0,1,1,1,2,4,2,3,1.458,6.79,-1.7,3.26,2.56,-3.55,-8.18,4.38,...



ภาพที่ 2-5 สายรหัสแทนต้นทางข้อมูลเข้าและลักษณะประจำ จำนวนชั้นซ่อน จำนวนนิวรอนในชั้นซ่อนและตัวถ่วงน้ำหนักการเชื่อมโยงของข่ายงานระบบประสาท

2.4.2 ถอดรหัสโครโมโซมของแต่ละสมาชิกในประชากรเพื่อให้ได้ตัวแปรตัดสินใจ

ตัวอย่างเช่น population : 1,0,0,1,1,1,0,1,1,1,2,4,2,3,1.458,6.79,-1.7,3.26,2.56,-3.55,-8.18,4.38,... ในตัวอย่างนี้ $n = 10$ โดยรหัสที่ใช้การ train คือ ยีนตัวที่ 1,4,5,6,8,9 และ 10 โดยมี Hidden layer จำนวน 2 ชั้น โดยชั้นที่ 1 มี 4 Hidden node ,ชั้นที่ 2 มี 2 Hidden node และชั้นที่ 3 มี 3 Hidden node

2.4.3 คำนวณค่าจุดประสงค์ (Objective Value)

คำนวณค่าจุดประสงค์ของแต่ละผลเฉลยโดยใช้ตัวแปรตัดสินใจที่ได้เป็นข้อมูลเข้า (Input) ของฟังก์ชันจุดประสงค์ ค่าจุดประสงค์ที่ได้คือ ค่า Mean of Sum of Square Errors ที่ได้จากสมการที่ 2-14

2.4.4 การประเมินค่าความแข็งแรง (Fitness Evaluation)

การประเมินค่าความแข็งแรงของสมาชิกแต่ละตัวโดยใช้ค่าจุดประสงค์สามารถคำนวณได้จาก

$$f_i = C_{\max} - J_i \quad (2-15)$$

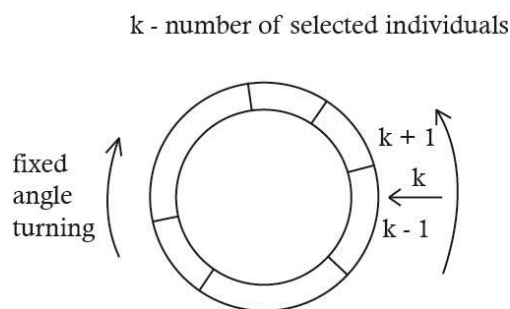
โดยที่ f_i คือ ค่าความแข็งแรงของสมาชิกตัวที่ i ในประชากร J_i คือ ค่าจุดประสงค์ซึ่งเป็นผลจากฟังก์ชันค่าใช้จ่ายในกรณีของสมาชิกตัวที่ i และ C_{\max} คือ ค่าที่เป็นบวกค่าหนึ่งซึ่งมีค่ามากกว่าค่าจุดประสงค์ที่มากที่สุดที่เป็นไปได้ โดยในที่นี้ได้ใช้ C_{\max} = จำนวน Output node

2.4.5 กลยุทธ์คัดแยกส่วนที่ดีที่สุด (Elitist Strategy)

เป็นกลยุทธ์หรือกฎฮิวริสติกที่ใช้ในขั้นตอนวิธีเชิงพันธุกรรมเพื่อเพิ่มโอกาสในการอยู่รอดของสมาชิกของประชากรที่ดีที่สุดหรือแข็งแรงที่สุด โดยเป็นการส่งผ่านสมาชิกที่มีค่าความแข็งแรงที่ดีที่สุดจำนวน n ไปยังรุ่นถัดไป

2.4.6 สร้างประชากรใหม่จากประชากรปัจจุบันโดยใช้ค่าความแข็งแรงในการกำหนดว่าควรเลือกสมาชิกตัวไหนจากประชากรปัจจุบันมาไว้ในประชากรใหม่

การเลือกสมาชิกของประชากรที่อยู่รอดจากรุ่นหนึ่งไปยังรุ่นถัดไป โดยใช้ค่าความแข็งแรงเป็นตัวกำหนดในการสร้างประชากรใหม่ในรุ่นถัดไป เทคนิคที่ใช้ในที่นี้คือ การคัดเลือกโดยการชักตัวอย่างทุกตัวแบบสุ่ม (Stochastic Universal Sampling Selection) (Baker, 1989) วิธีนี้จะใช้วงล้อรูเล็ตที่เอนเอียง (Biased Roulette Wheel) หรือวงล้อรูเล็ตที่มีช่องไม่เท่ากัน โดยช่องของสมาชิกที่มีค่าความแข็งแรงมากก็จะมีขนาดช่องที่ใหญ่กว่าสมาชิกที่มีค่าความแข็งแรงน้อย ทำให้สมาชิกที่มีค่าความแข็งแรงมากกว่ามีโอกาสที่จะถูกเลือกให้เป็นประชากรในรุ่นถัดไป จากนั้นจะมีการกำหนดตำแหน่งเริ่มต้นของตัวชี้โดยการสุ่ม เมื่อตำแหน่งของตัวชี้ตรงกับช่องใด สมาชิกในช่องนั้นก็จะถูกเลือกให้เป็นประชากรรุ่นถัดไป จากนั้นวงล้อรูเล็ตจะถูกหมุนไปเป็นมุม $360^\circ/n$ โดยที่ n คือจำนวนของสมาชิกในประชากร แผนภาพการคัดเลือกโดยการชักตัวอย่างทุกตัวแบบสุ่มได้แสดงไว้ในภาพที่ 2-6



ภาพที่ 2-6 การคัดเลือกโดยการชักตัวอย่างทุกตัวแบบเป็นสุ่ม (Stochastic universal sampling selection)

2.4.7 การปรับมาตราความแข็งแรง (Fitness Scaling)

เป็นการปรับระดับของความแตกต่างของค่าความแข็งแรง ซึ่งช่วงแรกในประชากรจะมีการกระจายค่าความแข็งแรงในระดับที่สูงมาก ทำให้สมาชิกที่มีความแข็งแรงสูงมากในช่วงแรกมีโอกาสถูกคัดเลือกเข้าสู่รุ่นถัดไปเป็นจำนวนมาก ส่งผลให้ความหลากหลายของยีนในประชากรมีการลดลงอย่างรวดเร็ว กำหนดให้ f เป็นค่าความแข็งแรงก่อนปรับมาตรา (Raw Fitness) และ f' เป็นค่าความแข็งแรงหลังปรับมาตรา (Scaled Fitness) ความสัมพันธ์ระหว่าง f กับ f' จะสามารถเขียนในรูปของความสัมพันธ์เชิงเส้นได้ดังนี้

$$f' = af + b \quad (2-16)$$

โดยที่ค่าสัมประสิทธิ์ (Coefficient) a และ b คือ ค่าที่ได้จากการคำนวณได้โดยอาศัยข้อกำหนดดังนี้

2.4.7.1 ค่าความแข็งแรงหลังการปรับมาตราเฉลี่ย (Average Scaled Fitness) หรือ f'_{avg}

จะต้องเท่ากับค่าความแข็งแรงก่อนการปรับมาตราเฉลี่ย (Average Raw Fitness) หรือ f_{avg}

$$f'_{avg} = f_{avg} \quad (2-17)$$

2.4.7.2 เพื่อเป็นการควบคุมอัตราการเลือกที่สมาชิกที่แข็งแรงที่สุดจะถูกเลือก ค่าความแข็งแรงหลังการปรับมาตราสูงสุด (Maximum Scaled Fitness) จะถูกกำหนดให้เป็น

$$f'_{max} = C_{mult} f_{avg} \quad (2-18)$$

โดยที่ C_{mult} เป็นอัตราการถูกเลือกที่ต้องการสำหรับสมาชิกที่แข็งแรงที่สุด สำหรับขั้นตอนวิธีเชิงพันธุกรรมที่มีขนาดของประชากรในช่วง 50-100 ค่าของ C_{mult} ที่เหมาะสมคือ 1.2-2.0

2.4.7.3 ในช่วงแรกของขั้นตอนวิธีเชิงพันธุกรรม f'_{\min} จะมีค่ามากกว่าศูนย์ ทำให้สามารถใช้วิธีข้างต้นได้ แต่เมื่อเข้าสู่ช่วงหลังของขั้นตอนวิธีเชิงพันธุกรรมการปรับมาตราความแข็งแรงตามวิธีข้างต้นจะทำให้ f'_{\min} จะมีค่าน้อยกว่าศูนย์ จึงสามารถแก้ไขได้โดยกำหนดให้ค่าความแข็งแรงหลังการปรับมาตราเฉลี่ยยังคงเท่ากับค่าความแข็งแรงก่อนปรับมาตราเฉลี่ย และกำหนดให้ค่าความแข็งแรงหลังการปรับมาตราต่ำสุด (Minimum Scaled Fitness) หรือ f'_{\min} เท่ากับศูนย์ การคำนวณค่าสัมประสิทธิ์ a และ b สำหรับการปรับมาตราเชิงเส้น (Linear Scaling) ตามรหัสเทียมที่แสดงในภาพที่ 2-6

```
if (f_min > (C_mult * f_avg - f_max) / C_mult - 1.0)) {
    delta = f_max - f_avg;
    a = (C_mult - 1.0) * f_avg / delta;
    b = f_avg * (f_max - C_mult * f_avg) / delta;
} else {
    delta = f_avg - f_min;
    a = f_avg / delta;
    b = -f_min * f_avg / delta;
}
```

ภาพที่ 2-7 รหัสเทียมสำหรับการคำนวณค่าสัมประสิทธิ์ที่ใช้ในการปรับมาตราความแข็งแรง

2.4.8 ทำการแปลงสมาชิกในประชากรใหม่โดยการครอสโอเวอร์ (Crossover) และการกลายพันธุ์ (Mutation)

2.4.8.1 การครอสโอเวอร์ (Crossover)

การสุ่มเลือกสมาชิกรุ่นพ่อแม่ (Parent Individual) จำนวน 2 ตัวจากประชากร เพื่อสร้างสมาชิกรุ่นลูก (Offspring Individual) ซึ่งจะมีจำนวน 2 ตัวเช่นกัน โอกาสที่จะมีการครอสโอเวอร์เกิดขึ้นระหว่างสมาชิกรุ่นพ่อแม่สองตัวที่สุ่มขึ้นมาจะถูกกำหนดโดยความน่าจะเป็นในการครอสโอเวอร์ (Crossover Probability) โดยปกติความน่าจะเป็นในการครอสโอเวอร์จะมีความอยู่ระหว่าง 0.7 ถึง 0.9

ในที่นี้ได้เลือกใช้การครอสโอเวอร์แบบ การครอสโอเวอร์เลขคณิต (Arithmetical Crossover) (Michalewicz and Janikow, 1991) โดย s_1 และ s_2 เป็นประชากรที่ประกอบจำนวนจริง ซึ่งแทนตัวถ่วงน้ำหนักการเชื่อมโยง ผลลัพธ์ที่ได้จากการครอสโอเวอร์เลขคณิตคือ

$$s'_1 = [rs_1 + (1-r)s_2] \quad (2-19)$$

และ

$$s'_2 = [rs_2 + (1-r)s_1] \quad (2-20)$$

2.4.8.2 การกลายพันธุ์ (Mutation)

ใช้ในการสร้างสมาชิกใหม่จากสมาชิกเดิมที่มีอยู่ โอกาสที่จะเกิดการกลายพันธุ์ในแต่ละยีนบนตัวโครโมโซมจะถูกกำหนดโดยความน่าจะเป็นในการกลายพันธุ์ (Mutation Probability) ตามปกติแล้วค่าความน่าจะเป็นในการกลายพันธุ์จะอยู่ระหว่าง 0 ถึง 0.1

ในที่นี้ได้เลือกใช้การกลายพันธุ์แบบ การกลายพันธุ์ขอบ (Boundary Mutation) เป็นการกลายพันธุ์ที่ทำให้เกิดการเปลี่ยนรหัสเพียงรหัสเดียวบนประชากร โดยรหัสที่เปลี่ยนไปจะเปลี่ยนเป็นค่า max หรือ min ซึ่งมีความน่าจะเป็นในการเปลี่ยนเป็น max หรือ min เท่ากัน ค่า max และ ค่า min ได้ถูกกำหนดไว้ในหัวข้อที่ 2.1

2.5 10-fold Cross Validation

10-fold Cross Validation หรือการแบ่งข้อมูลเป็น 10 ส่วนเท่า ๆ กันแล้วทำการทดสอบจำนวน 10 ส่วนทบ โดยแต่ละส่วนทบจะเปลี่ยนข้อมูลที่ใช้ในการทดสอบตั้งแต่ชุดที่ 1 ถึง 10 และใช้ข้อมูลที่เหลืออีก 9 ชุดในการเทรน โดยจะนำเอาค่าจุดประสงค์ที่ได้จากข้อมูลชุดทดสอบมาหาค่าเฉลี่ย (Ron Kohavi, 1995)

บทที่ 3

การทำงานและรายละเอียดของชุดข้อมูลที่ใช้ในการทดลอง

3.1 การทำงานของโปรแกรม

ขั้นตอนการทำงานของโปรแกรมแสดงไว้ในภาพที่ 3-1 เริ่มต้นจากการใช้ขั้นตอนวิธีเชิงพันธุกรรมในการฝึกสอนข่ายงานระบบประสาทโดยฟังก์ชันจุดประสงค์และฟังก์ชันความแข็งแรง (Fitness Function) เมื่อทำการคัดเลือก การครอสโอเวอร์เลขคณิต และการกลายพันธุ์ขอบ แล้วจึงฝึกสอนข่ายงานระบบประสาทโดยใช้ขั้นตอนวิธีแพร่ย้อนหลังและ 10-fold cross validation โดยมีจุดประสงค์เพื่อลดค่าจากฟังก์ชันค่าใช้จ่ายตามสมการที่ (2-14) และจะได้ 1 individual จากแต่ละส่วนทบ แล้วจึงสุ่ม 1 individual ออกมาเป็นตัวแทนของ individual ทั้งหมด จำนวนครั้งของการทำซ้ำ (Iteration) ขั้นตอนวิธีแพร่ย้อนหลังขึ้นกับผู้ใช้งาน เมื่อเสร็จจากขั้นตอนวิธีแพร่ย้อนหลังแล้วจึงจะกลับไปใช้ขั้นตอนวิธีเชิงพันธุกรรมกับข่ายงานระบบประสาทในรุ่นใหม่ที่ได้อีก

Algorithm Hybrid Genetic-Gradient Search(**P**)

```

//Input: Population of initial neural networks P

//Output: Population of evolved neural networks P and their
objective values O

calculate objective values O of neural networks P

for  $i \leftarrow 1$  to MaxGeneration do
    calculate fitness values F of neural networks P
    scale fitness values F of neural networks P
    select parent neural networks  $P_{Parent}$  from P based on F
    crossover  $P_{Parent}$  to obtain offspring neural networks  $P_{Offspring}$ 
    mutate  $P_{Offspring}$  to obtain mutated neural networks  $P_{Mutated}$ 
    for  $j \leftarrow 1$  to PopSize do
         $P[j] \leftarrow P_{Mutated}[j]$ 
    10-fold cross validation neural networks P using a back-
    propagation algorithm
    calculate objective values O of neural networks P
return P and O

```

ภาพที่ 3-1 รหัสเทียมสำหรับการคำนวณค่าสัมประสิทธิ์ที่ใช้ในการปรับมาตรการความแข็งแรง

3.2 ข้อมูลที่ใช้ในการทดลอง

3.2.1 Penetrance model

Penetrance model คือ ตารางที่แสดงตัวเลขของค่าความน่าจะเป็นที่ SNP ในช่องนั้นจะส่งผลให้เกิดโรค

3.2.2 โปรแกรมจีโนมซิม (genomeSIM)

ในการทดลองนี้ได้สร้างชุดข้อมูลที่ได้ในการทดลองจากโปรแกรมจีโนมซิม (genomeSIM) ซึ่งเป็นงานวิจัยของ Dudek et. al. (2006) โดยเป็นโปรแกรมที่ใช้ในการสร้างชุดข้อมูลของประชากรที่จะมีสถานะของการเป็นโรคหรือไม่เป็นโรคและชุดข้อมูลที่ได้จากโปรแกรมจะอยู่ในรูปของสายรหัส

3.2.3 รูปแบบปัญหาที่ใช้ในการทดลอง

ชุดข้อมูลที่ใช้ในการทดลองนี้เป็นรูปแบบปัญหาที่ใช้ในการทดลองด้านพันธุศาสตร์ (Genetic) ซึ่งมีทั้งหมด 12 ปัญหา (Becker et. al., 2005; Knapp et. al., 1994)

3.2.3.1 ตารางที่ 3-1 Penetrance ของปัญหา Epistasis 1

		Locus 2			
		0	1	2	
Locus 1	0	0	0	0	$\phi = 0.707$
	1	0	ϕ	ϕ	
	2	0	ϕ	ϕ	

3.2.3.2 ตารางที่ 3-2 Penetrance ของปัญหา Epistasis 2

		Locus 2			
		0	1	2	
Locus 1	0	0	ϕ	ϕ	$\phi = 0.778$
	1	0	0	0	
	2	0	0	0	

3.2.3.3 ตารางที่ 3-3 Penetrance ของปัญหา Epistasis 3

		Locus 2			
		0	1	2	
Locus 1	0	ϕ	0	0	$\phi = 0.900$
	1	0	0	0	
	2	0	0	0	

3.2.3.4 ตารางที่ 3-4 Penetrance ของปัญหา Epistasis 4

		Locus 2			
		0	1	2	
Locus 1	0	0	0	ϕ	$\phi = 0.911$
	1	0	0	ϕ	
	2	0	ϕ	ϕ	

3.2.3.5 ตารางที่ 3-5 Penetrance ของปัญหา Epistasis 5

		Locus 2			
		0	1	2	
Locus 1	0	0	0	0	$\phi = 0.799$
	1	0	0	ϕ	
	2	0	ϕ	ϕ	

3.2.3.6 ตารางที่ 3-6 Penetrance ของปัญหา Epistasis 6

		Locus 2			
		0	1	2	
Locus 1	0	0	0	ϕ	$\phi = 1.000$
	1	0	0	ϕ	
	2	ϕ	ϕ	0	

3.2.3.7 ตารางที่ 3-7 Penetrance ของปัญหา Heterogeneity 1

		Locus 2			
		0	1	2	
Locus 1	0	0	ϕ	ϕ	$\phi = 0.495$
	1	ϕ	ψ	ψ	
	2	ϕ	ψ	ψ	

$\psi = 2\phi - \phi^2$

3.2.3.8 ตารางที่ 3-8 Penetrance ของปัญหา Heterogeneity 2

		Locus 2			
		0	1	2	
Locus 1	0	0	ϕ	ϕ	$\phi = 0.660$
	1	0	ϕ	ϕ	
	2	ϕ	ψ	ψ	

$\psi = 2\phi - \phi^2$

3.2.3.9 ตารางที่ 3-9 Penetrance ของปัญหา Heterogeneity 3

		Locus 2			
		0	1	2	
Locus 1	0	0	0	ϕ	$\phi = 1.000$
	1	0	0	ϕ	
	2	ϕ	ϕ	ψ	$\psi = 2\phi - \phi^2$

3.2.3.10 ตารางที่ 3-10 Penetrance ของปัญหา S1

		Locus 2			
		0	1	2	
Locus 1	0	0	ϕ	ϕ	$\phi = 0.522$
	1	ϕ	ϕ	ϕ	
	2	ϕ	ϕ	ϕ	

3.2.3.11 ตารางที่ 3-11 Penetrance ของปัญหา S2

		Locus 2			
		0	1	2	
Locus 1	0	0	ϕ	ϕ	$\phi = 0.574$
	1	0	ϕ	ϕ	
	2	1	1	1	

3.2.3.12 ตารางที่ 3-12 Penetrance ของปัญหา S3

Locus 2				
		0	1	2
Locus 1	0	0	0	ϕ
	1	0	ϕ	1
	2	ϕ	1	1

$\phi = 0.512$

3.2.4 ตัวอย่างข้อมูล

ลักษณะชุดข้อมูลที่ได้จากโปรแกรมจีโนมซิมจะอยู่ในรูปของสายรหัส ในตอนแรก โปรแกรมจีโนมซิมสร้างชุดข้อมูลออกมาโดยที่อักขรตัวแรกจะแทนสถานะของการเป็นโรค หรือไม่โรค โดยถ้ามีค่าเป็น 1 หมายถึงสถานะเป็นโรค ส่วนถ้ามีค่าเป็น 0 หมายถึงสถานะไม่เป็นโรค ดังภาพที่ 4-13

1	1	0	0	1	1	0	2	1	0	1
1	1	0	2	2	1	1	2	0	1	1
0	0	1	1	1	1	0	0	2	1	1
1	1	1	1	1	1	0	0	0	1	0
0	0	0	1	1	0	1	0	0	2	2
0	1	1	1	1	0	0	0	1	1	2
1	1	0	2	2	1	0	2	0	1	1
0	0	0	0	0	0	1	0	0	1	0
1	1	2	0	1	1	1	0	2	2	1
0	0	2	1	1	1	0	1	2	2	1

ภาพที่ 4-13 ตัวอย่างข้อมูลในตอนแรก

แต่เพื่อให้งานต่อการนำไปใช้งาน จึงได้มีการจัดกลุ่มชุดข้อมูลใหม่และเปลี่ยนตำแหน่งอักขรที่แสดงสถานะของการเป็นโรคหรือไม่เป็นโรคให้เป็นอักขรตัวสุดท้ายแทน ชุดข้อมูลที่ใช้ใน

การทดลองได้แสดงไว้ในภาพที่ 4-14 และภาพที่ 4-15 เป็นตัวอย่างของแฟ้มข้อมูลตาราง Penetrance ที่ใช้ในโปรแกรมจีโนมซิม

1	0	0	1	1	0	2	1	0	1	1
1	0	2	2	1	0	2	0	1	1	1
1	2	0	1	1	1	0	2	2	1	1
1	0	2	2	1	1	2	0	1	1	1
1	1	1	1	1	0	0	0	1	0	1
0	0	1	1	0	1	0	0	2	2	0
0	1	1	1	1	0	0	2	1	1	0
1	1	1	1	0	0	0	1	1	2	0
0	0	0	0	0	1	0	0	1	0	0
0	2	1	1	1	0	1	2	2	1	0

ภาพที่ 4-14 ตัวอย่างข้อมูลหลังจากทำการเรียงใหม่

```
# any comments can be like this
# loci need to start with letter A and so on

# list of disease loci in genome (start with 1)
DISEASELOCI 1 5

# penetrance values
# all penetrances must follow the keyword
# first locus should be A followed by B etc.
PENTABL
AABB 0
AABb 0
AAbb 0
AaBB 0
AaBb 0.707
Aabb 0.707
aaBB 0
aaBb 0.707
aabb 0.707
```

ภาพที่ 4-15 ตัวอย่างแฟ้มข้อมูลตาราง Penetrance

3.2.5 เป้าหมาย

เป้าหมายของระบบคือสามารถค้นหา SNP ที่เป็นต้นเหตุที่ทำให้เกิดโรค

บทที่ 4

ผลการทดลองและการวิเคราะห์ผลการทดลอง

4.1 การทดลอง

การทำงานของโปรแกรมใน 1 ปี จะทำงานตามรหัสเทียมในภาพที่ 3-1 จำนวน 100 ครั้ง ในแต่ละครั้งจะได้โครงสร้างของข่ายงานระบบประสาท 1 โครงสร้าง ซึ่งสามารถบอกได้ว่า SNP ตัวใดมีผลต่อการเกิดโรค เมื่อทำครบ 100 ครั้งจะหาว่าใน 100 ครั้ง SNP แต่ละตัวมีผลต่อการเกิดโรคกี่ครั้งและทำการหาอัตราความสำเร็จ (Success Rate) อัตราความสำเร็จ (Success Rate) คือ อัตราระหว่างจำนวนครั้งของ SNP ตัวที่สนใจส่งผลให้เกิดโรคเมื่อเปรียบเทียบกับจำนวนครั้งในการทำงานของโปรแกรม โดยนับจำนวนครั้งที่เจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค และจำนวนครั้งที่เจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค

4.2 ผลการทดลอง

ผลการทดลองจาก 12 ปัญหา โดยให้มีทั้งหมด 10 SNPs และ SNP ตัวที่ถูกกำหนดให้ส่งผลต่อการเกิดโรคคือ SNP ตัวที่ 1 และ 5

4.1.1 ผลการทดลองจากปัญหา Epistasis 1

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-16 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-16

```

loop : 90, fitness : 0.9631257688629711
poppulation : 1,0,1,0,1,0,0,1,0,1,1,3,2,2,3.4641070739,3.6516154327...

loop : 91, fitness : 0.973122925907812
poppulation : 1,1,1,1,1,1,1,1,1,2,3,3,3,-0.09739737347,16.2860025...

loop : 92, fitness : 0.9697204430888139
poppulation : 1,1,1,1,1,1,1,1,1,2,3,3,3,-5.87603600,16.526718058...

loop : 93, fitness : 0.9666275664056815
poppulation : 1,0,0,1,1,1,0,0,0,2,3,3,3,5.6033320479,4.1844035641...

loop : 94, fitness : 0.9653050433887589
poppulation : 1,0,0,1,1,0,0,0,0,2,3,3,3,6.3607178618,2.7137464302...

loop : 95, fitness : 0.9658049719612933
poppulation : 1,0,0,0,1,1,1,0,1,0,2,3,3,3,-4.2311583469,4.871705953...

loop : 96, fitness : 0.9643740522132384
poppulation : 1,0,1,0,1,0,0,1,0,0,2,3,3,3,-6.7750687351,1.913738792...

loop : 97, fitness : 0.9667320979434598
poppulation : 1,0,0,0,1,0,0,0,0,2,2,3,3,2.77598968495,-1.92043404...

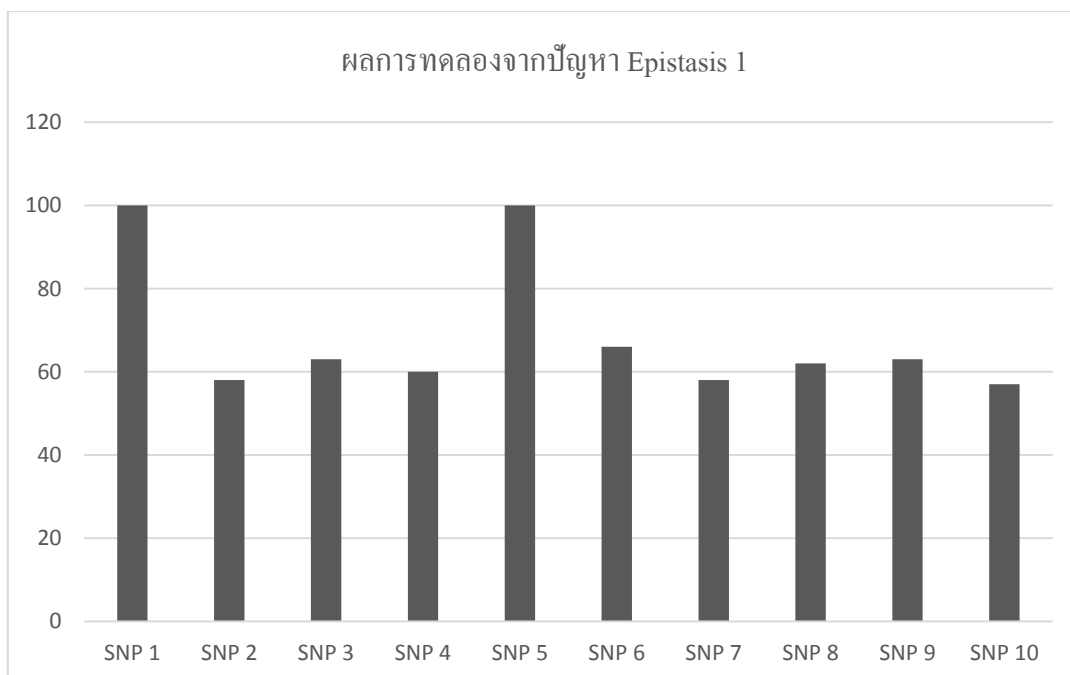
loop : 98, fitness : 0.9660003450730614
poppulation : 1,1,1,0,1,1,0,0,1,0,2,3,3,3,6.6557871946,0.6445631775...

loop : 99, fitness : 0.9663946012249205
poppulation : 1,1,1,1,1,1,1,1,1,2,4,3,3,14.358655250,1.9063281984...

1) 100 2) 58 3) 63 4) 60 5) 100 6) 66 7) 58 8) 62 9) 63 10) 57
Runtime : 0 days, 9 hours, 59 minutes, 4 seconds.

```

ภาพที่ 4-16 ผลการทดลองจากปัญหา Epistasis 1



ภาพที่ 4-17 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 1

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.2 ผลการทดลองจากปัญหา Epistasis 2

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-18 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-19

```

loop : 90, fitness : 0.9428717924585802
poppulation : 1,0,0,1,1,0,0,0,0,1,2,3,3,3,3.0236703759,-4.067021250...

loop : 91, fitness : 0.9438221876801545
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,6.6341597940,3.8027236338...

loop : 92, fitness : 0.9430603173353966
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,-0.819910953,-3.651352542...

loop : 93, fitness : 0.9438082721347675
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,4,4.5608349458,8.905345916...

loop : 94, fitness : 0.9450235115811405
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,5.0202203388,4.9789422483...

loop : 95, fitness : 0.9441781898520898
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,3,10.759562666,0.8036224257...

loop : 96, fitness : 0.9442371381917093
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,-0.2670124021,8.738213075...

loop : 97, fitness : 0.941456866510824
poppulation : 1,0,0,0,1,0,1,0,0,0,1,2,2,4,5.5393511750,2.8152713530...

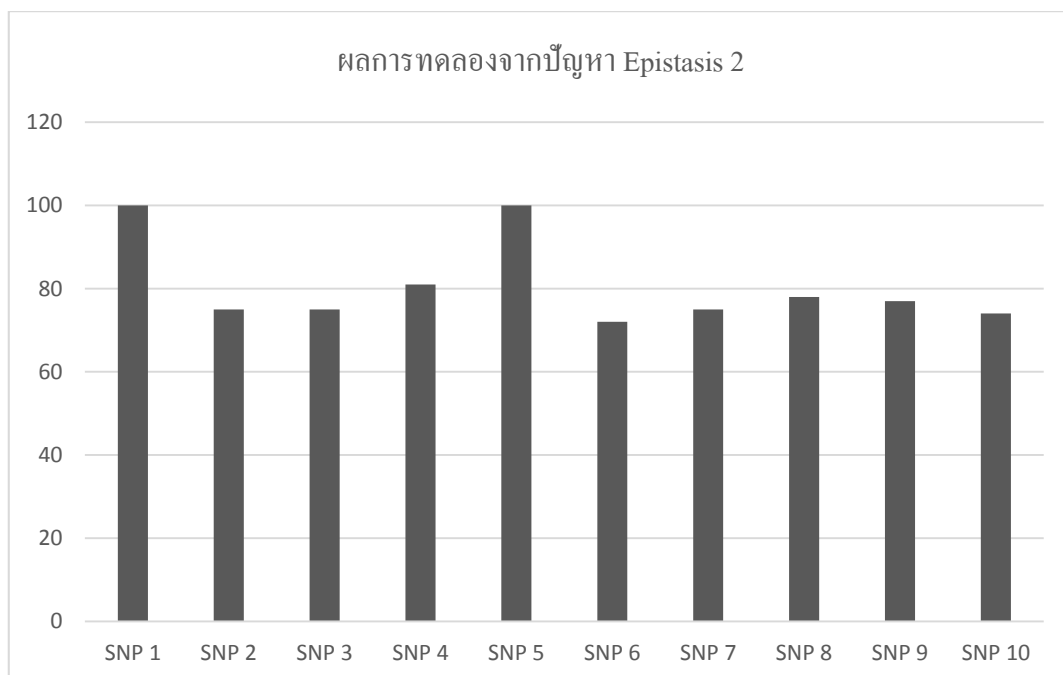
loop : 98, fitness : 0.9442641954890006
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,-3.9858099642,-1.92521318...

loop : 99, fitness : 0.9476086911378454
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,-3.84471706E-5,-3.157306...

1) 100 2) 75 3) 75 4) 81 5) 100 6) 72 7) 75 8) 78 9) 77 10) 74
Runtime : 0 days, 10 hours, 55 minutes, 8 seconds

```

ภาพที่ 4-18 ผลการทดลองจากปัญหา Epistasis 2



ภาพที่ 4-19 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 2

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.3 ผลการทดลองจากปัญหา Epistasis 3

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-20 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-21

```

loop : 90, fitness : 0.9925158066825008
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,4,4,6.078763503,5.94369458...

loop : 91, fitness : 0.9925857396256524
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,1.18897046517,6.92452131...

loop : 92, fitness : 0.9926230694792121
poppulation : 1,0,0,0,1,1,0,0,0,0,2,3,3,3,2.346139767,0.3391034907...

loop : 93, fitness : 0.994065953742783
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,3,-1.9751323642,3.79613142...

loop : 94, fitness : 0.9930064007464106
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,6.24779597855,6.17888376...

loop : 95, fitness : 0.9925618469355962
poppulation : 1,0,0,0,1,0,0,0,1,0,2,3,3,1,3.9820398912,6.0013402332...

loop : 96, fitness : 0.9926247103649539
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,3,3.6514642787,6.3022249961...

loop : 97, fitness : 0.9925430620442339
poppulation : 1,1,0,0,1,1,1,1,1,0,2,3,3,3,1.572249492,-3.2202043822...

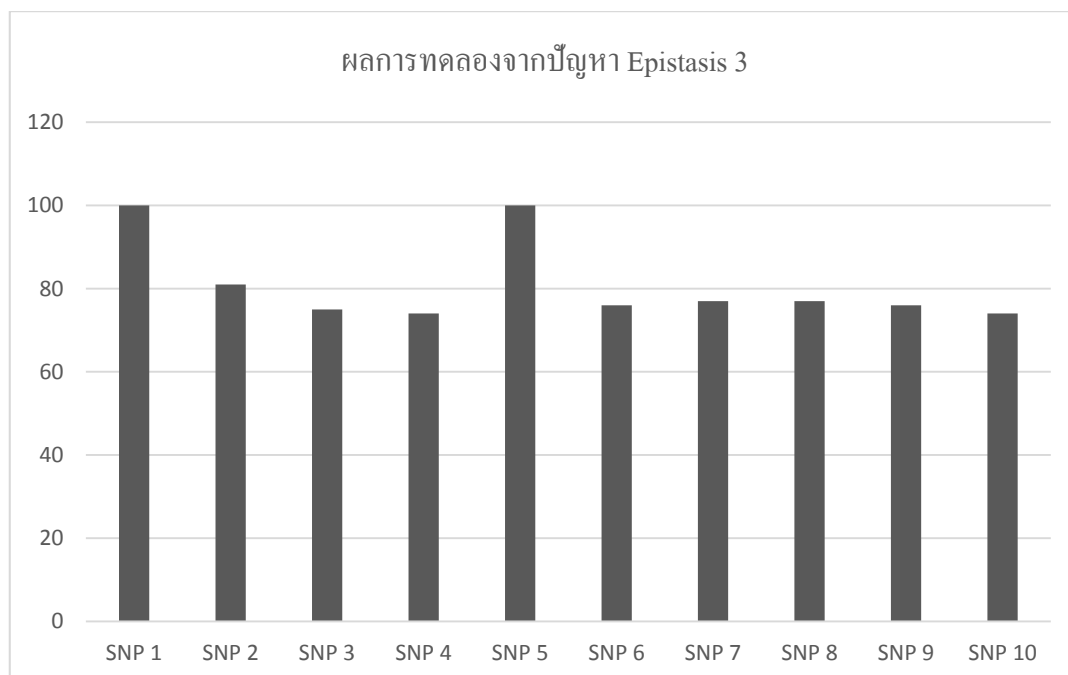
loop : 98, fitness : 0.9925396086924471
poppulation : 1,1,1,0,1,1,1,1,1,1,2,3,3,3,4.3622377642,3.0465327459...

loop : 99, fitness : 0.9926510865699493
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,1.004024493,5.1675700048...

1) 100 2) 81 3) 75 4) 74 5) 100 6) 76 7) 77 8) 77 9) 76 10) 74
Runtime : 0 days, 10 hours, 44 minutes, 33 seconds.

```

ภาพที่ 4-20 ผลการทดลองจากปัญหา Epistasis 3



ภาพที่ 4-21 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 3

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.4 ผลการทดลองจากปัญหา Epistasis 4

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-22 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-23

```

loop : 90, fitness : 0.9922713920064676
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,2,16.985803377,4.0000001019...

loop : 91, fitness : 0.9901229358502132
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,13.999638826,11.999924866...

loop : 92, fitness : 0.99008703582977
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,7.9998797427,3.9998959248...

loop : 93, fitness : 0.9874099131198341
poppulation : 1,0,0,1,1,1,0,0,0,0,2,3,3,3,18.731696926,14.869379309...

loop : 94, fitness : 0.9878076798778571
poppulation : 1,0,1,0,1,0,0,0,0,0,2,3,3,3,21.758902768,13.990138780...

loop : 95, fitness : 0.989995374864521
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,1.99992580493,3.999619263...

loop : 96, fitness : 0.9899709108878079
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,4,7.0244079486,10.022883858...

loop : 97, fitness : 0.9905136578111827
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,29.6263460926,8.999465009...

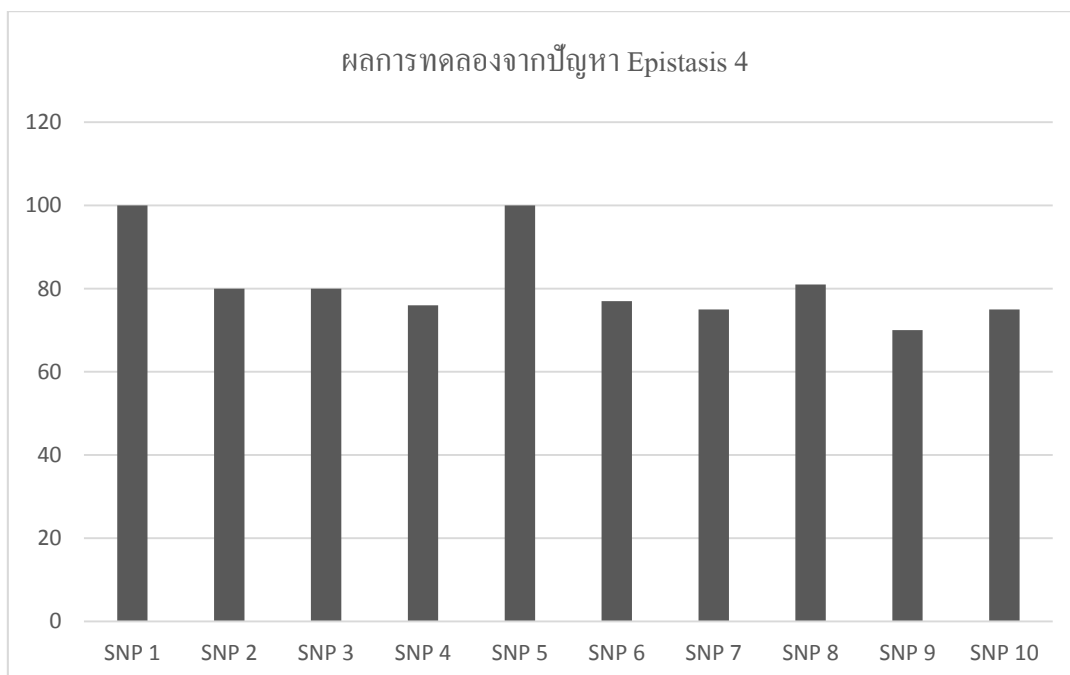
loop : 98, fitness : 0.9896443872972536
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,2.9945555686,36.670103301...

loop : 99, fitness : 0.9899987191141761
poppulation : 1,1,0,0,1,1,1,0,0,0,2,3,3,3,-34.2496479,-22.342053262...

1) 100 2) 80 3) 80 4) 76 5) 100 6) 77 7) 75 8) 81 9) 70 10) 75
Runtime : 0 days, 10 hours, 21 minutes, 52 seconds.

```

ภาพที่ 4-22 ผลการทดลองจากปัญหา Epistasis 4



ภาพที่ 4-23 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 4

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.5 ผลการทดลองจากปัญหา Epistasis 5

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-24 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-25

```

loop : 90, fitness : 0.9949228676175216
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,1.99999913469,6.9999503216...

loop : 91, fitness : 0.9974709780284012
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,10.054526492,15.011072886...

loop : 92, fitness : 0.9924662420042991
poppulation : 1,1,1,1,1,1,1,1,1,1,3,4,3,3,12.5221685862,38.58531895...

loop : 93, fitness : 0.9940485235634067
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,3,-6.0000000115,5.968575547...

loop : 94, fitness : 0.9943422644688146
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,2,6.00218188182,1.999999982...

loop : 95, fitness : 0.9927489161878956
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,0.9874107452,-3.000018521...

loop : 96, fitness : 0.9963943883114937
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,2.9999968446,3.9999987779...

loop : 97, fitness : 0.996799203238239
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,25.5462447132,10.996480512...

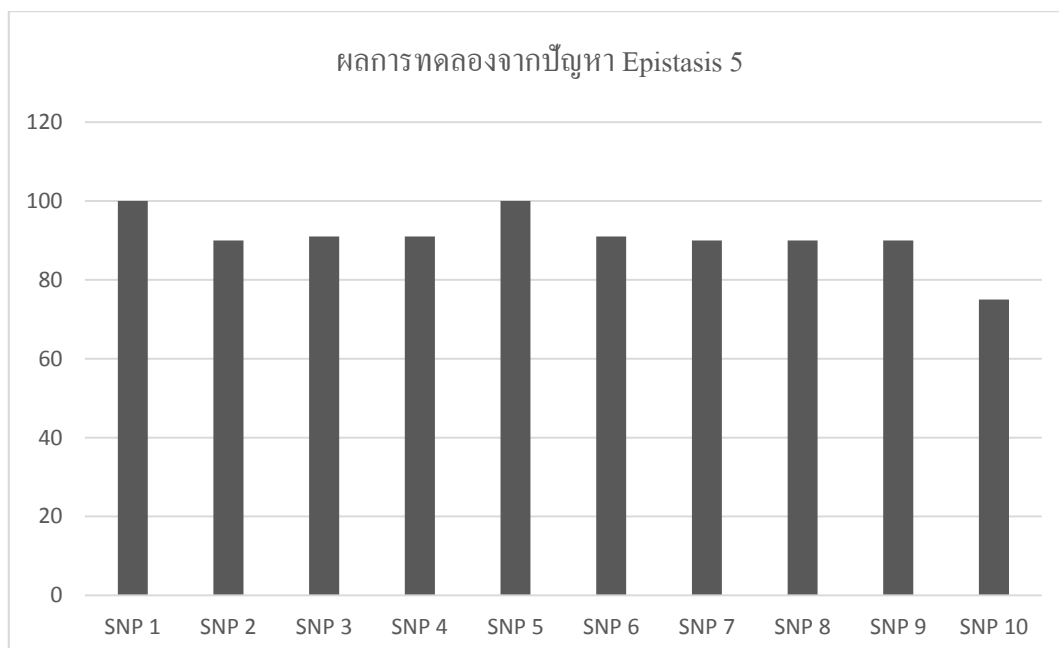
loop : 98, fitness : 0.9925896429049504
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,3,22.138789069,20.1126324991...

loop : 99, fitness : 0.9963881978508068
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,19.8975013018,11.96103694...

1) 100 2) 90 3) 91 4) 91 5) 100 6) 91 7) 90 8) 90 9) 90 10) 91
Runtime : 0 days, 10 hours, 37 minutes, 52 seconds.

```

ภาพที่ 4-24 ผลการทดลองจากปัญหา Epistasis 5



ภาพที่ 4-25 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 5

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.6 ผลการทดลองจากปัญหา Epistasis 6

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-26 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-27

```

loop : 90, fitness : 0.9967331860801735
population : 1,1,1,1,1,1,1,1,1,2,3,3,3,-14.878424409,-3.04286405...

loop : 91, fitness : 0.9974436156614345
population : 1,0,0,0,1,0,0,0,0,1,2,3,3,3,12.78086616857,4.33454116...

loop : 92, fitness : 0.9970732014806672
population : 1,0,0,1,1,0,0,0,0,0,2,3,3,3,-0.34455827011,8.44148987...

loop : 93, fitness : 0.9955006648732894
population : 1,1,0,1,1,0,0,0,0,0,2,3,3,3,10.696769184,2.6475667059...

loop : 94, fitness : 0.9975014270246895
population : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,6.99998982966,18.05904603...

loop : 95, fitness : 0.9984942496308918
population : 1,0,0,0,1,0,0,0,0,0,2,3,3,3,-3.2251500026,10.37691602...

loop : 96, fitness : 0.99992000388194
population : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,-6.1496161027,-7.0637856...

loop : 97, fitness : 0.9970263484844611
population : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,0.9756219533,5.7402962620...

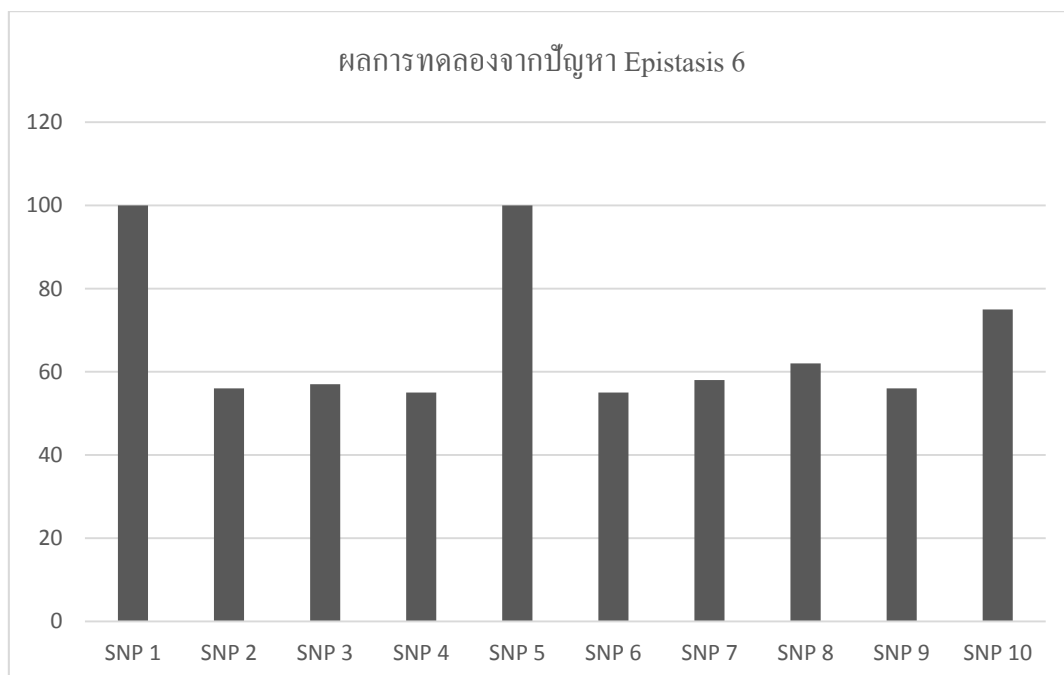
loop : 98, fitness : 0.9967600211749322
population : 1,1,0,0,1,0,0,0,0,0,2,3,3,3,9.41813585,-1.9881228965...

loop : 99, fitness : 0.9899585091836673
population : 1,0,1,0,1,0,1,1,0,1,2,4,2,4,8.4034617357,11.160780381...

1) 100 2) 56 3) 57 4) 55 5) 100 6) 55 7) 58 8) 62 9) 56 10) 63
Runtime : 0 days, 10 hours, 20 minutes, 22 seconds.

```

ภาพที่ 4-26 ผลการทดลองจากปัญหา Epistasis 6



ภาพที่ 4-27 กราฟแสดงอัตราความสำเร็จของปัญหา Epistasis 6

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.7 ผลการทดลองจากปัญหา Heterogeneity 1

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-28 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-29

```

loop : 90, fitness : 0.9354913969559451
poppulation : 1,1,1,1,1,1,1,0,1,1,2,3,3,3,-1.5701856763,-5.96769893...

loop : 91, fitness : 0.937569773990551
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,6.0223988805,16.184250267...

loop : 92, fitness : 0.9354879735404134
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,17.004946114,2.7467347994...

loop : 93, fitness : 0.9357299788254764
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,-2.463704668,2.523660086...

loop : 94, fitness : 0.9354366769923957
poppulation : 1,1,1,0,1,0,0,1,1,0,2,3,3,3,7.5484188877,4.0067427041...

loop : 95, fitness : 0.935705115676551
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,3,-4.14840892,-2.186484460...

loop : 96, fitness : 0.9351071190562731
poppulation : 1,0,0,0,1,1,0,0,0,0,1,3,3,3,-3.6989045302,4.54455201...

loop : 97, fitness : 0.9376629736259652
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,5.0509530382,0.9124202009...

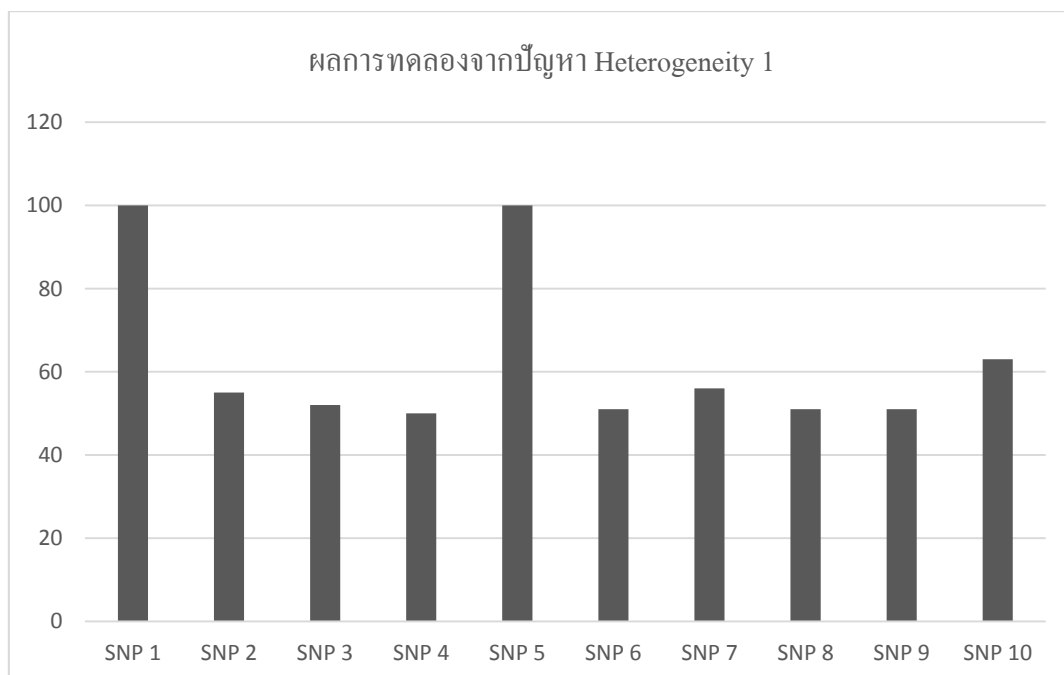
loop : 98, fitness : 0.9360977031234929
poppulation : 1,0,0,0,1,0,0,0,1,0,2,3,3,3,-1.1993847378,4.000560538...

loop : 99, fitness : 0.9341166397408692
poppulation : 1,0,0,0,1,1,0,0,0,0,1,2,1,3,3.25404206105,5.9275929410...

1) 100 2) 55 3) 52 4) 50 5) 100 6) 51 7) 56 8) 51 9) 51 10) 56
Runtime : 0 days, 9 hours, 41 minutes, 44 seconds.

```

ภาพที่ 4-28 ผลการทดลองจากปัญหา Heterogeneity 1



ภาพที่ 4-29 กราฟแสดงอัตราความสำเร็จของปัญหา Heterogeneity 1

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.1.8 ผลการทดลองจากปัญหา Heterogeneity 2

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-30 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-31

```

loop : 90, fitness : 0.8880573336700003
poppulation : 1,0,0,0,0,0,0,0,0,0,2,3,3,3,-1.0000154143,-6.38413381...

loop : 91, fitness : 0.888024304424539
poppulation : 1,0,0,0,1,0,0,0,0,0,1,3,3,3,2.9853742271,-2.773067376...

loop : 92, fitness : 0.8883544031977941
poppulation : 1,0,0,0,0,0,0,0,1,0,1,2,3,4,-0.995484336,-2.116489121...

loop : 93, fitness : 0.8870427019210578
poppulation : 1,0,1,1,0,1,1,0,0,1,2,3,3,3,3.8316994460,4.02305794125...

loop : 94, fitness : 0.8879147495550974
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,2.9287668293,1.0498183562...

loop : 95, fitness : 0.8853473609912912
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,2.78380922725,1.958749617...

loop : 96, fitness : 0.8841070268992026
poppulation : 1,0,0,0,0,0,0,0,0,0,1,1,3,3,3,-0.9423863331,0.994490861...

loop : 97, fitness : 0.8880322320672359
poppulation : 1,0,0,0,0,0,0,0,0,0,2,3,3,2,2.000036975,-5.1656455207...

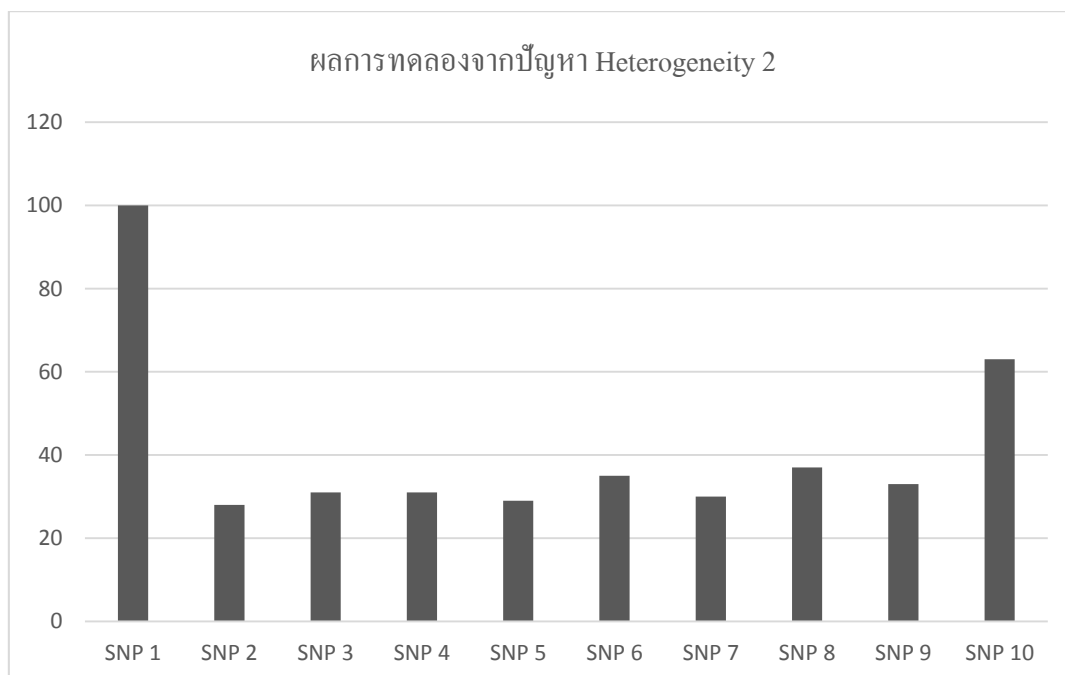
loop : 98, fitness : 0.8881513331375793
poppulation : 1,0,0,0,0,0,0,0,0,0,2,3,3,3,8.3964441453,0.7321333125...

loop : 99, fitness : 0.886247118207966
poppulation : 1,0,0,0,0,1,1,1,0,1,1,3,3,4,7.3858149767,3.7996038498...

1) 100  2) 28  3) 31  4) 31  5) 29  6) 35  7) 30  8) 37  9) 33  10) 33
Runtime: 0 days, 10 hours, 35 minutes, 50 seconds.

```

ภาพที่ 4-30 ผลการทดลองจากปัญหา Heterogeneity 2



ภาพที่ 4-31 กราฟแสดงอัตราความสำเร็จของปัญหา Heterogeneity 2

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 29 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.2.9 ผลการทดลองจากปัญหา Heterogeneity 3

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-32 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-33

```

loop : 90, fitness : 0.9985742444057221
poppulation : 1,1,0,1,1,0,0,0,0,0,2,3,3,3,-7.007443752,-9.584662884...

loop : 91, fitness : 0.9943739649270498
poppulation : 1,1,0,1,1,1,1,0,0,1,2,3,3,3,-15.698656932,-1.38333840...

loop : 92, fitness : 0.9935836346516346
poppulation : 1,1,1,1,1,1,1,1,0,1,2,4,3,3,7.048007406,6.6721624428...

loop : 93, fitness : 0.9999965031257974
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,4,5.999362015,11.433416448...

loop : 94, fitness : 0.9989647412282678
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,2,19.82578717892,14.5366282...

loop : 95, fitness : 0.9999901242456438
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,15.994544373,16.06336536...

loop : 96, fitness : 0.9993208640942177
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,10.77427023,10.050065486...

loop : 97, fitness : 0.999958977953666
poppulation : 1,0,0,1,1,0,0,0,0,0,2,3,3,3,5.5844430370,16.667043241...

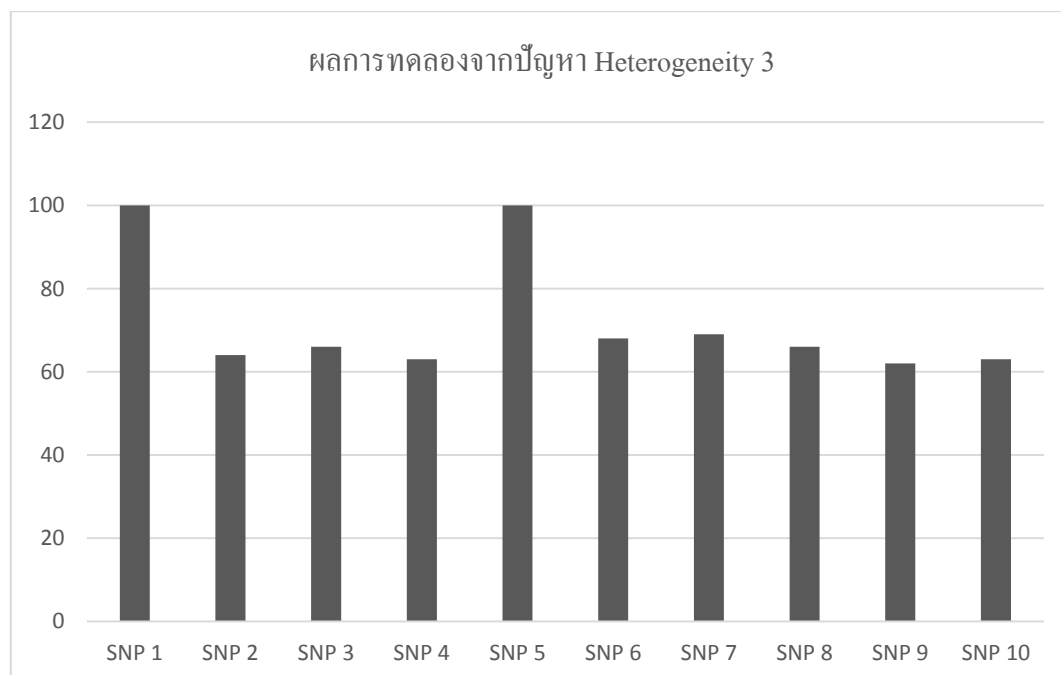
loop : 98, fitness : 0.9999979833884846
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,23.0002531932,5.999991037...

loop : 99, fitness : 0.9983205169133679
poppulation : 1,0,0,1,1,0,0,1,0,1,1,3,2,4,7.1307610918,6.508055336...

1) 100 2) 64 3) 66 4) 63 5) 100 6) 68 7) 69 8) 66 9) 62 10) 63
Runtime : 0 days, 10 hours, 35 minutes, 20 seconds.

```

ภาพที่ 4-32 ผลการทดลองจากปัญหา Heterogeneity 3



ภาพที่ 4-33 กราฟแสดงอัตราความสำเร็จของปัญหา Heterogeneity 3

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.2.10 ผลการทดลองจากปัญหา S1

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-34 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-35

```

loop : 90, fitness : 0.9363609428565874
poppulation : 1,0,0,0,1,0,1,0,1,0,1,3,3,3,-2.078474923,-4.963384132...

loop : 91, fitness : 0.9386043305786504
poppulation : 1,0,0,0,1,0,0,0,0,0,1,3,3,3,5.9810060008,2.6578478849...

loop : 92, fitness : 0.9404719887518432
poppulation : 1,1,1,1,1,1,1,1,1,2,3,3,4,2.24971022436,6.514096898...

loop : 93, fitness : 0.9386157265947779
poppulation : 1,0,0,0,1,1,0,0,0,0,2,3,2,3,8.2820525121,2.8389138248...

loop : 94, fitness : 0.9381470000099723
poppulation : 1,1,1,1,1,1,1,1,1,2,4,4,4,2.6521369016,0.9995921177...

loop : 95, fitness : 0.9376433916720965
poppulation : 1,0,0,0,1,1,1,0,0,0,1,3,5,3,3.9832415574,2.0813319914...

loop : 96, fitness : 0.938896484422379
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,3,-3.0815443912,2.810350971...

loop : 97, fitness : 0.9377130565477236
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,3,2.6199333564,3.1673039165...

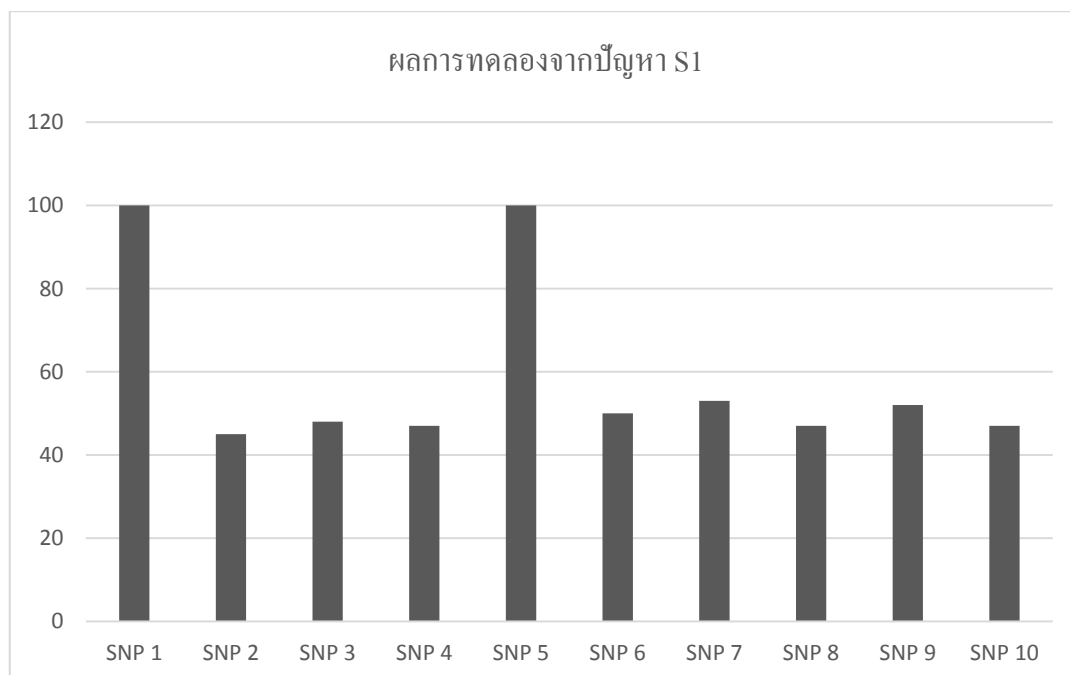
loop : 98, fitness : 0.9389362029419211
poppulation : 1,1,1,1,1,1,1,1,1,2,3,3,3,3.2893467495,7.4802703336...

loop : 99, fitness : 0.9414309877601654
poppulation : 1,1,1,1,1,1,1,1,1,2,3,3,3,6.8126923909,1.1070373549...

1) 100 2) 45 3) 48 4) 47 5) 100 6) 50 7) 53 8) 47 9) 52 10) 47
Runtime : 0 days, 10 hours, 15 minutes, 48 seconds.

```

ภาพที่ 4-34 ผลการทดลองจากปัญหา S1



ภาพที่ 4-35 กราฟแสดงอัตราความสำเร็จของปัญหา S1

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.2.11 ผลการทดลองจากปัญหา S2

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-36 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-37

```

loop : 90, fitness : 0.8984576171705376
poppulation : 1,0,0,1,0,0,0,0,0,1,2,3,3,3,-4.174032984,-0.585358236...

loop : 91, fitness : 0.9040797353013809
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,2,4.3695184920,6.36488807150...

loop : 92, fitness : 0.900309391071224
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,3,5.0419804878,5.8539695866...

loop : 93, fitness : 0.8997802072451447
poppulation : 1,0,0,0,0,0,0,0,0,0,1,2,2,2,1.9383612674,-2.085274032...

loop : 94, fitness : 0.8986356638108621
poppulation : 1,0,0,0,0,0,0,0,0,0,2,3,3,3,0.56901381304,4.932149325...

loop : 95, fitness : 0.8997864860947608
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,6.2471081898,5.5441703889...

loop : 96, fitness : 0.8990704470110089
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,4,4.0278914576,5.6890683187...

loop : 97, fitness : 0.8996988467163578
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,5.2543421532,3.6026937490...

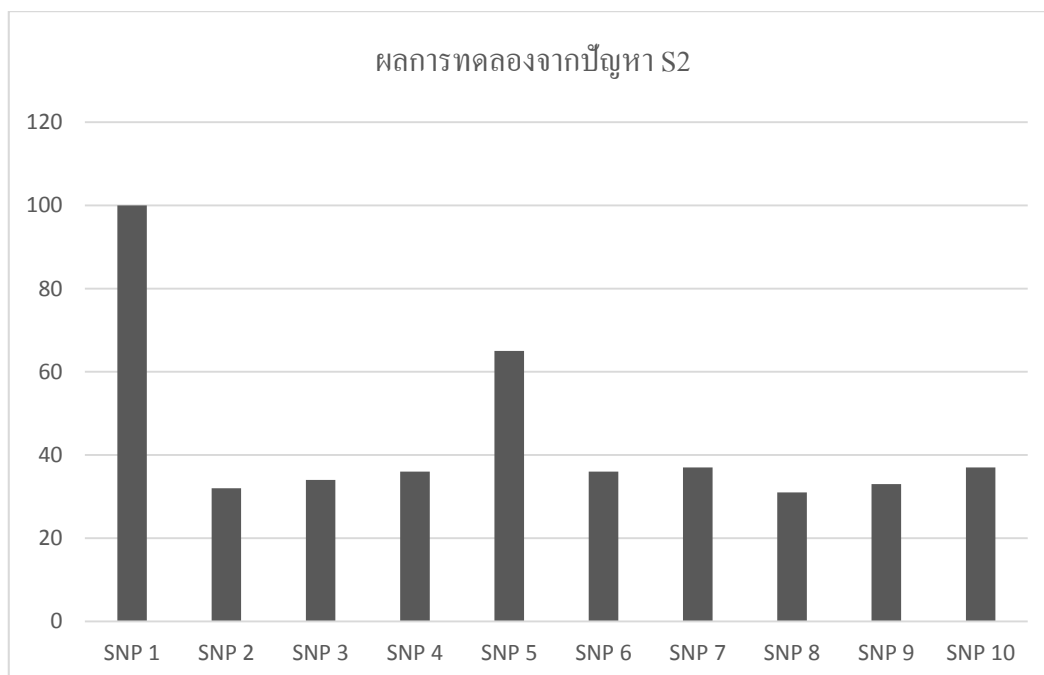
loop : 98, fitness : 0.8980135016807207
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,4,3,8.1726936962,4.4635194333...

loop : 99, fitness : 0.8994668433294174
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,6.0785084432,5.8516324995...

1) 100 2) 32 3) 34 4) 36 5) 65 6) 36 7) 37 8) 31 9) 33 10) 37
Runtime : 0 days, 10 hours, 29 minutes, 15 seconds.

```

ภาพที่ 4-36 ผลการทดลองจากปัญหา S2



ภาพที่ 4-37 กราฟแสดงอัตราความสำเร็จของปัญหา S2

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 65 ครั้ง
- กรณีเจออย่างน้อย 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.2.12 ผลการทดลองจากปัญหา S3

ผลการทดลองจากการทำงาน 10 ครั้งสุดท้ายแสดงไว้ในภาพที่ 4-38 และกราฟแสดงอัตราความสำเร็จแสดงไว้ในภาพที่ 4-39

```

loop : 90, fitness : 0.9670670459800217
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,3,15.395244502,11.843093754...

loop : 91, fitness : 0.9667731519315658
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,-20.289261306,1.971462886...

loop : 92, fitness : 0.9669853015460327
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,4,3,30.237449917,20.035857851...

loop : 93, fitness : 0.9668193892364011
poppulation : 1,1,1,1,1,1,1,1,1,1,2,4,3,4,15.999909766,14.028785767...

loop : 94, fitness : 0.9668887145476011
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,-19.840606358,4.998593564...

loop : 95, fitness : 0.96670895180421
poppulation : 1,0,0,0,1,0,0,0,1,0,2,3,3,3,20.869738007,16.356991679...

loop : 96, fitness : 0.9665575347917188
poppulation : 1,1,0,1,1,1,1,1,1,1,2,3,3,3,4.040818071,-12.949322427...

loop : 97, fitness : 0.9607333173939647
poppulation : 1,1,1,1,1,1,1,1,1,1,2,3,3,3,14.406037806,3.9457617310...

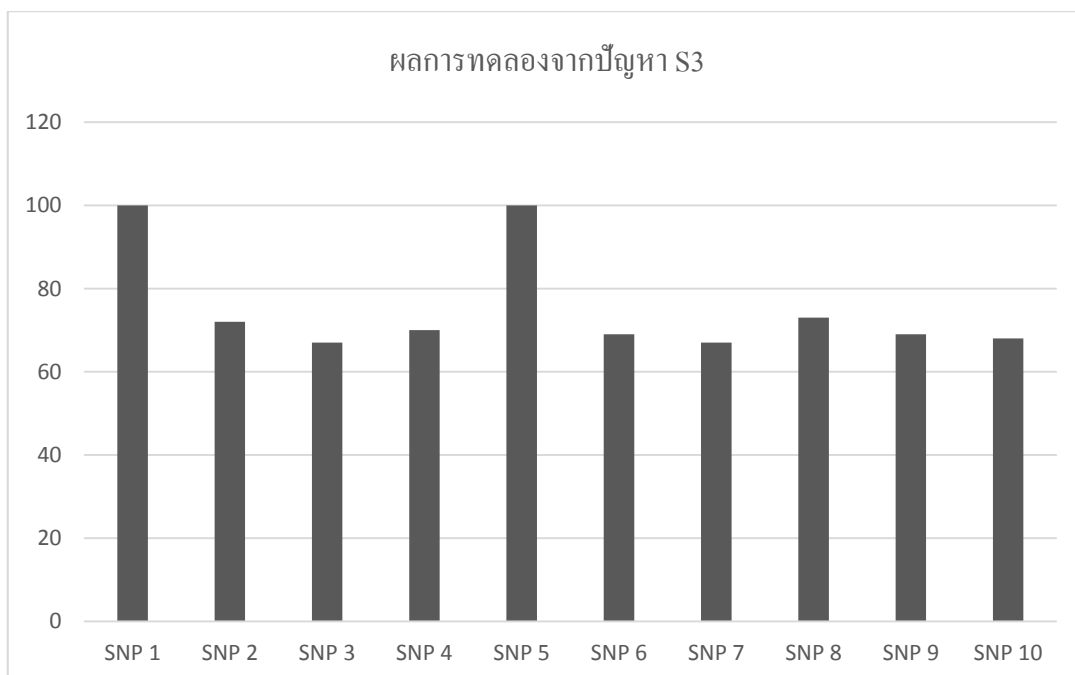
loop : 98, fitness : 0.9667419146776697
poppulation : 1,0,0,1,1,0,1,0,0,0,2,3,3,2,-0.525783526,0.1132792423...

loop : 99, fitness : 0.9667322382886508
poppulation : 1,0,0,0,1,0,0,0,0,0,2,3,3,3,13.935750589,16.5284969832...

1) 100 2) 72 3) 67 4) 70 5) 100 6) 69 7) 67 8) 73 9) 69 10) 68
Runtime : 0 days, 9 hours, 59 minutes, 6 seconds.

```

ภาพที่ 4-38 ผลการทดลองจากปัญหา S3



ภาพที่ 4-39 กราฟแสดงอัตราความสำเร็จของปัญหา S3

อัตราความสำเร็จ (Success Rate)

- กรณีเจอทั้ง 2 SNPs ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง
- กรณีเจอน้อยกว่า 1 SNP ที่ถูกกำหนดให้ส่งผลต่อการเกิดโรค = 100 ครั้ง

4.3 ตารางเวลาในการทดลอง

ตารางเวลาการทดลองของทั้ง 12 ปัญหาบน Server Intel(R) Xeon(R) X3330 2.66GHz quad-core processor and 4GB of RAM จะได้ระยะเวลาที่ใช้ในการทดลองตามที่แสดงในตารางต่อไปนี้

ตารางที่ 4-1 เวลาในการทดลองของทั้ง 12 ปัญหา

ปัญหา	เวลาที่ใช้ในการทดลอง
Epistasis 1	9 hours 59 minutes 4 seconds
Epistasis 2	10 hours 55 minutes 8 seconds
Epistasis 3	10 hours 44 minutes 33 seconds
Epistasis 4	10 hours 21 minutes 52 seconds
Epistasis 5	10 hours 37 minutes 52 seconds
Epistasis 6	10 hours 20 minutes 22 seconds
Heterogeneity 1	9 hours 41 minutes 44 seconds
Heterogeneity 2	10 hours 35 minutes 50 seconds
Heterogeneity 3	10 hours 35 minutes 20 seconds
S1	10 hours 15 minutes 48 seconds
S2	10 hours 29 minutes 15 seconds
S3	9 hours 59 minutes 6 seconds

บทที่ 5

สรุปผลการวิจัย

ในปริญญานิพนธ์ฉบับนี้ได้นำเสนอการตรวจจับความสัมพันธ์เชิงพันธุกรรมโดยทำงานระบบประสาทเชิงวิวัฒนาการ เพื่อทดสอบความสามารถของแบบจำลองในการตรวจหา SNP ที่ส่งผลให้เกิดโรค โดยทำการทดลองกับปัญหาทั้งหมด 12 รูปแบบ

ในการทดลองพบว่าผลจากการทดลองซ้ำ 100 ครั้งในแต่ละปัญหาทั้ง 12 ปัญหา พบว่าสามารถตรวจพบ SNP ที่ส่งผลให้เกิดโรคครบทั้งสองตำแหน่ง 10 ปัญหา คิดเป็นร้อยละ 83.33 และตรวจพบ SNP ที่ส่งผลให้เกิดโรคเพียงตำแหน่งเดียว 2 ปัญหา คิดเป็นร้อยละ 16.67

จากการทดลองนี้พบว่าการตรวจจับความสัมพันธ์เชิงพันธุกรรมโดยทำงานระบบประสาทเชิงวิวัฒนาการนั้นมีความแม่นยำในการวิเคราะห์และตรวจหา SNP ที่ส่งผลให้เกิดโรค สำหรับแนวทางในการพัฒนาแนวทางแรกคือการให้จำนวน attribute ที่ส่งผลต่อการเกิดโรคมีผลต่อค่าความแข็งแรง อาจทำได้โดยการนำค่าความแข็งแรงที่ได้คูณด้วยอัตราส่วนระหว่างจำนวน attribute ที่ไม่ถูกนำมาใช้ในการฝึกสอนกับจำนวน attribute ทั้งหมด และแนวทางที่สองคือการใช้ชุดข้อมูลหลาย ๆ ชุดต่อ 1 ปัญหาเพื่อหาค่า statistical power

เอกสารอ้างอิง

1. Aleksander, I. and Morton, H. (1990). An introduction to neural computing. London, UK: Chapman and Hall.
2. Baker, J. E. (1989). An analysis of the effects of selection in genetic algorithms Ph.D. Thesis, Computer Science Department, Vanderbilt University, Nashville, TN.
3. Becker, T., Schumacher, J., Cichon, S., Baur, M.P., Knapp, M.: Haplotype Interaction Analysis of Unlinked Regions. *Genetic Epidemiology*. 29 (2005), 313–322.
4. Dudek, S.M., Motsinger, A.A., Velez, D.R., Williams, S.M., Ritchie, M.D.: Data Simulation Software for Whole-Genome Association and Other Studies in Human Genetics. In: Altman, R.B., Dunker, A.K., Hunter, L., Murray, T., Klein, T.E. (eds.): *Pacific Symposium on Biocomputing 2006*. World Scientific, Singapore (2006), 499–510.
5. Goldberg, D. E. (1989). *Genetic algorithms: In search, optimization and machine learning*. Reading, MA: Addison-Wesley.
6. Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York, NY: Macmillan.
7. Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
8. Kawato, M., Uno, Y., Isobe, M. and Suzuki, R. (1988). Hierarchical neural network model for voluntary movement with application to robotics. *IEE Control Systems Magazine*, 8(2), 8-16.
9. Knapp, M., Seuchter, S.A., Baur, M.P.: Two-Locus Disease Models with Two MarkerLoci: The Power of Affected-Sib-Pair Tests. *American Journal of Human Genetics* 55 (1994) 1030–1041
10. Michalewicz, Z. and Janikow, C. Z. (1991). Genetic algorithms for numerical optimization. *Statistics and Computing*, 1(2), 75–91.

เอกสารอ้างอิง (ต่อ)

11. Ron Kohavi: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. The International Joint Conference on Artificial Intelligence (IJCAI) (1995)
12. Rumelhart, D. E. and McClelland, J. L. (Eds.) (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1. Cambridge, MA: MIT Press.

ประวัติผู้แต่ง

ปริญญานิพนธ์เรื่อง : การตรวจจับความสัมพันธ์เชิงพันธุกรรมโดยข่ายงานระบบ
ประสาทเชิงวิวัฒนาการ

สาขาวิชา : วิศวกรรมคอมพิวเตอร์

ภาควิชา : วิศวกรรมไฟฟ้าและคอมพิวเตอร์

คณะ : วิศวกรรมศาสตร์

ชื่อ : นางสาวพรเพ็ญ อาชีวะเกษะ

ประวัติ

เกิดเมื่อวันที่ 5 มกราคม พ.ศ. 2536 อยู่บ้านเลขที่ 42/574 หมู่บ้านเขาวพรรณ ถนนบางกรวย-ไทรน้อย ตำบลบางสีทอง อำเภอบางกรวย จังหวัดนนทบุรี สำเร็จการศึกษาจากโรงเรียนสตรี นนทบุรี จังหวัดนนทบุรี สาขาวิทยาศาสตร์-คณิตศาสตร์ ปีการศึกษา 2553 และสำเร็จการศึกษาในระดับปริญญาตรี สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ปีการศึกษา 2557

ชื่อ : นางสาวอรุณรัตน์ อมะลัยเชียร

ประวัติ

เกิดเมื่อวันที่ 14 มีนาคม พ.ศ. 2536 อยู่บ้านเลขที่ 90/71 หมู่ 1 ถนนบางกรวย-ไทรน้อย ตำบลบางกรวย อำเภอบางกรวย จังหวัดนนทบุรี สำเร็จการศึกษาระดับประกาศนียบัตรวิชาชีพ สาขาไฟฟ้าและอิเล็กทรอนิกส์ โครงการภาษาอังกฤษ จากโรงเรียนเตรียมวิศวกรรมศาสตร์ วิทยาลัยเทคโนโลยีอุตสาหกรรม ปีการศึกษา 2553 และสำเร็จการศึกษาในระดับปริญญาตรี สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ปีการศึกษา 2557