การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม

PROTOTYPE SELECTION BASED ON MINIMAL CONSISTENT SUBSET AND GENETIC ALGORITHMS

มารุต หงษ์สามารถ
MARUT HONGSAMART

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2552
KMITL - 2009 - EN - M - 070 - 039

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม

PROTOTYPE SELECTION BASED ON MINIMAL CONSISTENT SUBSET AND GENETIC ALGORITHMS

T105133

มารุต หงษ์สามารถ

MARUT HONGSAMART

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

วน
ม 484ก
2552

เลขหมู่.....
เลขทะเบียน.....105133
วัน,เดือน,ปี.....16 พ.ย. 2552

b.12166157
i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2552

KMITL-2009-EN-M-070-039

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PROTOTYPE SELECTION BASED ON MINIMAL CONSISTENT SUBSET AND GENETIC ALGORITHMS

MARUT HONGSAMART

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN COMPUTER ENGINEERING

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2009

KMITL-2009-EN-M-070-039

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2009

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม

Thesis Title Prototype Selection based on Minimal Consistent Subset and Genetic Algorithms

นักศึกษา นายมารุต หงษ์สามารถ

รหัสประจำตัว 49060701

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.ดร.บุญธีร์ เครือตราชู

หมายเลขวิทยานิพนธ์ KMITL-2009-EN-M-070-039

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
ดร.วรวัฒน์	ลิ้มโภคา	
ผศ.ดร.สมศักดิ์	วลัยรัชต์	
รศ.ดร.ชม	กิ้มปาน	
รศ.กฤตวัน	ศิริบูรณ์	
รศ.ดร.บุญธีร์	เครือตราชู	

วัน / เดือน/ ปี ที่สอบ วันอังคารที่ 31 มีนาคม พ.ศ. 2552 เวลา 13.00-15.00 น.

สถานที่สอบ ณ อาคาร A ชั้น 3 ห้องประชุม 2

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิศวกรรมศาสตร์ รับรองแล้ว

(รองศาสตราจารย์ ดร.กอบชัย เดชหาญ)

คณบดี คณะวิศวกรรมศาสตร์

วันที่ 31 มีนาคม พ.ศ. 2552

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม
นักศึกษา	นายมารุต หงษ์สามารถ
รหัสนักศึกษา	49060701
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
พ.ศ.	2552
อาจารย์ที่ปรึกษา	รศ.ดร.บุญธีร์ เครือตราชู

บทคัดย่อ

การจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule เป็นที่ใช้กันอย่างแพร่หลาย แต่ในกรณีที่ข้อมูลอ้างอิงมีจำนวนมาก จะทำให้เกิดปัญหาการคำนวณในปริมาณสูง และปัญหาความต้องการหน่วยความจำจำนวนมาก จึงต้องมีการนำกระบวนการเลือกโปรโตไทป์มาทำการลดจำนวนข้อมูลอ้างอิงลงให้มีจำนวนน้อยที่สุด ในขณะที่ยังคงสามารถรักษาประสิทธิภาพการจำแนกให้ใกล้เคียงกับการใช้ข้อมูลเดิมอ้างอิงได้ วิทยานิพนธ์ฉบับนี้จึงเสนอการเลือกโปรโตไทป์เพื่อลดจำนวนข้อมูลอ้างอิงลง โดยนำวิธี Minimal Consistent Subset (MCS) และเจเนติกอัลกอริธึม (GAs) มาทำงานร่วมกัน โดยเรียกวิธีใหม่นี้เรียกว่า "Hybrid Genetic Algorithms (HGAs)" ซึ่งเป็นการนำข้อดีของแต่ละวิธีมาแก้ข้อด้อยของอีกวิธี โดยใช้ GAs ทำการค้นหาแบบกว้างๆ ก่อนจะใช้วิธี MCS กับคำตอบที่ได้นั้นอีกที อีกทั้งยังมีการเพิ่มเติมเทคนิค "Singleton" เพื่อลดขนาดของปริภูมิคำตอบ และเทคนิค "Diversity" เพื่อกระจายการค้นหาคำตอบให้กว้างขึ้น โดยทำการทดสอบ HGAs กับชุดข้อมูลมาตรฐานที่มีขนาดต่างๆ กัน ก็จะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่า MCS และ GAs ในทุกชุดข้อมูล โดยเฉพาะอย่างยิ่งในชุดข้อมูล IRIS ก็จะเห็นว่า HGAs สามารถให้คำตอบเป็นผลลัพธ์ที่ดีที่สุดทุกครั้งที่ทำการทดลอง ในขณะที่ GAs ก็สามารถหาได้ แต่เวลาที่ใช้ในการทำงานก็นานกว่าอย่างชัดเจน ส่วนวิธี MCS ไม่สามารถให้คำตอบที่เป็นผลลัพธ์ที่ดีที่สุดได้เลย ส่วนในชุดข้อมูลที่มีจำนวนข้อมูลมาก (ชุดข้อมูล YEAST, MUSK และ PEN) ก็จะเห็นว่าการทำงานของ GAs แย่ลงอย่างชัดเจน ในขณะที่ HGAs ยังสามารถให้ผลลัพธ์ที่ดีกว่าวิธี MCS เล็กน้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Prototype Selection based on Minimal Consistent Subset and Genetic Algorithms
Student	Mr. Marut Hongsamart
Student ID.	49060701
Degree	Master of Engineering
Program	Computer Engineering
Year	2009
Thesis Advisor	Assoc.Prof.Dr. Boontee Kruatrachue

Abstract

The nearest neighbor rule classification has been used widely. But in case of large amount of reference data, the nearest neighbor rule classification requires large memory space and expensive computation time. So, prototype selection process is needed to reduce the amount of reference data, while maintain the classification efficiency as similar as original reference data. This thesis proposes a condensing prototype selection algorithm by combining Minimal Consistent Subset (MCS) method with Genetic Algorithms, called "Hybrid Genetic Algorithms (HGAs)", which use the benefit of each method to compensate drawback of another method. In HGAs, GAs provide a global search while MCS provides a local search. Addition, "Singleton" and "Diversity" techniques are adopted to reduce search space and perform wider search respectively. HGAs had been tested on several-size standard datasets, which obtain better results than MCS and GAs in every dataset. Especially in IRIS, which HGAs obtain an optimal result in every experiment, while GAs can obtain an optimal result in some of the experiment but requiring more processing time, and MCS can't obtain an optimal result at all. On large-size dataset (YEAST, MUSK and PEN). GAs perform much worst, while HGAs have a little better result than MCS.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้จะความสำเร็จลุล่วงไปไม่ได้ หากไม่ได้รับความกรุณาเป็นอย่างสูงจาก อาจารย์ผู้ควบคุมวิทยานิพนธ์ คือ รองศาสตราจารย์ ดร.บุญธีร์ เครือตราชู ที่ช่วยให้คำปรึกษาเป็นอย่างดีเสมอมา อีกทั้งยังให้วิชาความรู้ในหัวข้อต่างๆ ที่เกี่ยวข้องกับงานวิจัยตั้งแต่เริ่มต้นจนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ต้องขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้ด้วย

ขอขอบพระคุณกรรมการสอบวิทยานิพนธ์ทุกท่านที่ได้กรุณาให้คำแนะนำต่างๆ จนทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์ สามารถครอบคลุมเนื้อหาของงานวิจัยในแง่มุมต่างๆ อย่างครบถ้วน

ขอขอบพระคุณอาจารย์ทุกท่าน ที่ได้ให้ความรู้ต่างๆ ทั้งโดยตรง และโดยอ้อม ตลอดจนนักวิจัยทุกท่านที่เอื้อเฟื้องานวิจัย จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ทุกคน ที่ช่วยเหลือสนับสนุนในเรื่องต่างๆ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณบัณฑิตศึกษา คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ให้การสนับสนุนในการทำวิจัย จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

สุดท้ายขอขอบพระคุณบิดา มารดา และครอบครัวของข้าพเจ้า ที่ได้ให้การสนับสนุนในเรื่องต่างๆ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

หวังว่าผู้อ่านคงได้จะความรู้จากวิทยานิพนธ์ฉบับนี้ไม่มากก็น้อย และหากวิทยานิพนธ์ฉบับนี้มีข้อผิดพลาดประการใด ข้าพเจ้าขอน้อมรับไว้แต่เพียงผู้เดียว

มารุต หงษ์สามารถ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง X

สารบัญรูป XIV

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 สมมติฐานของการศึกษา 2

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย 3

1.5 ขอบเขตการวิจัย 3

1.6 ขั้นตอนการศึกษา 4

1.7 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย 4

1.8 โครงสร้างของวิทยานิพนธ์ 5

บทที่ 2 นิยามและทฤษฎีพื้นฐานที่เกี่ยวข้อง 6

2.1 นิยามพื้นฐาน (Basic Definition) 6

2.1.1 ออบเจ็ค (Object) 6

2.1.2 ปริภูมิคุณลักษณะเฉพาะ (Feature Space) 7

2.1.3 ปริภูมิคำตอบ (Solution Space) 7

2.1.4 ชุดข้อมูลฝึกสอน (Training Dataset) 7

2.1.5 ชุดข้อมูลทดสอบ (Testing Dataset) 8

2.1.6 เซตชนิดของข้อมูล (Category Set) 8

2.1.7 โปรโตไทป์ผลลัพธ์ (Prototype Set) 8

2.1.8 ออบเจ็คไม่ทราบชนิด (Unknown Object) 8

2.1.9 ฟังก์ชั่นคำนวณระยะทาง (Distance Function) 9

2.2 การจำแนกข้อมูล (Classification) 9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.3 Nearest Neighbor Rule (NN Rule) 10

2.4 การวัดระยะทางแบบยูคลิด (Euclidean Distance) 12

2.5 การเลือกโปรโตไทป์ (Prototype Selection) 13

2.5.1 การเลือกโปรโตไทป์เพื่อลดจำนวนโปรโตไทป์ (Condensing Technique) 14

2.5.2 การเลือกโปรโตไทป์เพื่อกำจัดข้อมูลที่ไม่เข้าพวก (Editing Technique) 14

2.5.3 การเลือกโปรโตไทป์ที่ให้ผลลัพธ์ที่สร้างขึ้นใหม่ (Replacement Technique) 15

2.5.4 การเลือกโปรโตไทป์ที่ให้ผลลัพธ์เป็นเซตย่อย (Selection Technique) 15

2.6 ขอบเขตการตัดสินใจ (Decision Boundary) 16

2.7 คุณสมบัติความสอดคล้อง (Consistency Property) 17

บทที่ 3 งานวิจัยที่เกี่ยวข้อง 18

3.1 Minimal Consistent Set Identification (MCS) 18

3.1.1 นิยามที่ควรทราบเกี่ยวกับวิธี MCS 20

3.1.1.1 ออบเจ็คต่างชนิดที่อยู่ใกล้ที่สุด (Nearest Unlike Neighbor) 20

3.1.1.2 รัศมีขอบเขตต่างชนิด (NUN Boundary) 20

3.1.1.3 ความครอบคลุม (Cover) 21

3.1.1.4 ค่าความสามารถในการครอบคลุม (Cover Value) 22

3.1.1.5 เซตของออบเจ็คที่น่าจะเป็นโปรโตไทป์ (Candidate Set) 22

3.1.2 รายละเอียดขั้นตอนการทำงานของวิธี MCS 22

3.1.2.1 กระบวนการคำนวณระยะทาง (Distance Calculating Process) 22

3.1.2.2 กระบวนการคำนวณความครอบคลุม (Cover Calculating Process) 23

3.1.2.3 กระบวนการเลือกออบเจ็ค (Object Selecting Process) 28

3.2 Genetic Algorithms (GAs) 38

3.2.1 นิยามที่ควรทราบเกี่ยวกับ GAs 40

3.2.1.1 โครโมโซม (Chromosome) 40

3.2.1.2 ประชากร (Population) 41

3.2.1.3 จำนวนประชากร (N_{pop}) 41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.2.1.4 จำนวนเจเนอเรชั่น (N_{gen}) 41

3.2.1.5 ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) 41

3.2.1.6 ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) 41

3.2.1.7 ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) 41

3.2.1.8 ฟังก์ชั่นคำนวณค่าความเหมาะสม (Fitness Function) 41

3.2.1.9 โครโมโซมที่ดีที่สุด (The Best Chromosome) 42

3.2.2 รายละเอียดการทำงานของ GAs 43

3.2.2.1 กระบวนการกำหนดค่าเริ่มต้น (Initial Process) 44

3.2.2.2 กระบวนการครอสโอเวอร์ (Crossover Process) 47

3.2.2.3 กระบวนการกลายพันธุ์ (Mutation Process) 49

3.2.2.4 กระบวนการเลือกสรรประชากรรุ่นถัดไป (Selection Process) 50

บทที่ 4 การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม 53

4.1 ภาพรวมการทำงานของระบบ 54

4.2 รายละเอียดขั้นตอนการทำงาน 57

4.2.1 กระบวนการสร้างตารางระยะทาง (Distance Calculating Process) 57

4.2.2 กระบวนการกำหนดค่าเริ่มต้น (Initial Process) 57

4.2.2.1 การสร้างประชากรเริ่มต้น (Population Creation) 57

4.2.2.2 การสุ่มกำหนดค่าให้กับประชากรเริ่มต้น (Population Initialization) ..58

4.2.2.3 การสร้างโครโมโซมที่ดีที่สุด (The Best Chromosome Creation) 58

4.2.3 กระบวนการ MCS (MCS Process) 59

4.2.3.1 กระบวนการคำนวณความครอบคลุม (Cover Calculating Process) 59

4.2.3.2 กระบวนการการเลือกออบเจ็ค (Object Selecting Process) 60

4.2.4 กระบวนการครอสโอเวอร์ (Crossover Process) 63

4.2.4.1 การสร้างรูเล็ทวีล (Roulette Wheel Creation) 63

4.2.4.2 การสุ่มเลือกโครโมโซมพ่อ-แม่ (Parent Selection) 63

4.2.5 กระบวนการกลายพันธุ์ (Mutation Process) 64

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2.6 กระบวนการเลือกสรรประชากรรุ่นถัดไป (Selection Process) 64

4.2.6.1 การนำโครโมโซมทั้งหมดมารวมกัน (Chromosome Pooling) 64

4.2.6.2 การเลือกโครโมโซมรุ่นถัดไป (Chromosome Selection) 64

4.2.6.3 การปรับปรุงค่าโครโมโซมที่ดีที่สุด (The Best Chromosome Updating) 65

4.3 การเพิ่มประสิทธิภาพการทำงานด้วยเทคนิค “Singleton” 65

4.4 การเพิ่มประสิทธิภาพการทำงานด้วยเทคนิค “Diversity” 67

4.4.1 การวัดระยะทางแบบฮัมมิ่ง (Hamming Distance) 68

4.4.2 การประยุกต์ใช้หลักการ “Diversity” 68

4.4.2.1 การประยุกต์หลักการ “Diversity” ในกระบวนการครอสโอเวอร์ (Diversity Crossover) 69

4.4.2.2 การประยุกต์หลักการ “Diversity” ในกระบวนการเลือกสรรประชากรรุ่นถัดไป (Diversity Selection) 72

4.5 HGAs แบบต่างๆ 76

4.5.1 HGAs แบบของเทคนิคการเลือกของกระบวนการ MCS 77

4.5.2 HGAs แบบของเทคนิค “Singleton” 81

4.5.3 HGAs แบบของเทคนิค “Diversity” 82

บทที่ 5 การทดลองและผลการทดลอง 84

5.1 ชุดข้อมูลที่ใช้ในการทดลอง 85

5.1.1 ชุดข้อมูลขนาดเล็ก 85

5.1.1.1 ชุดข้อมูล IRIS 85

5.1.1.2 ชุดข้อมูล ECOLI 85

5.1.2 ชุดข้อมูลขนาดกลาง 86

5.1.2.1 ชุดข้อมูล CREDIT 86

5.1.2.2 ชุดข้อมูล YEAST 86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

5.1.3 ชุดข้อมูลขนาดใหญ่ 86
5.1.3.1 ชุดข้อมูล MUSK 86
5.1.3.2 ชุดข้อมูล PEN 86
5.2 ผลการทดลอง 87
5.2.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม 87
5.2.1.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล IRIS 88
5.2.1.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล ECOLI 89
5.2.1.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล CREDIT 90
5.2.1.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล YEAST 91
5.2.1.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล MUSK 92
5.2.1.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล PEN 93
5.2.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ 96

บทที่ 6 สรุปผลการทดลองและข้อเสนอแนะ 110
6.1 สรุปผลการทดลอง 110
6.2 ข้อเสนอแนะ 113

เอกสารอ้างอิง 114

ภาคผนวก 116

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

ภาคผนวก ก. รายละเอียดการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูลต่างๆ 117
ภาคผนวก ข. รายละเอียดการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูลต่างๆ 122
ภาคผนวก ค. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ 127
ภาคผนวก ง. ตัวอย่างการทดลองของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ 133
ภาคผนวก จ. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อย 136
ภาคผนวก ฉ. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” 140
ภาคผนวก ช. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” 144
ภาคผนวก ซ. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของลำดับการทำงาน 150
ภาคผนวก ฌ. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs 156
ภาคผนวก ซ. งานวิจัยที่ได้รับการตีพิมพ์ 162

ประวัติผู้เขียน 169

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 ข้อมูลระยะทางของชุดข้อมูลตัวอย่างที่ 2 23

3.2 ข้อมูลระยะทางของชุดข้อมูลตัวอย่างที่ 2 ที่เรียงลำดับแล้ว 24

3.3 การเลือก NUN ในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 25

3.4 ตารางความครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 27

3.5 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 27

3.6 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 เป็นโปรโตไทป์แล้ว 29

3.7 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 และ B2 เป็นโปรโตไทป์แล้ว 30

3.8 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1, B2 และ A4 เป็นโปรโตไทป์แล้ว 30

3.9 การเลือก NUN ในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 32

3.10 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 32

3.11 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 เป็นโปรโตไทป์แล้ว 33

3.12 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 และ B2 เป็นโปรโตไทป์แล้ว 34

3.13 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1, B2 และ A4 เป็นโปรโตไทป์แล้ว 34

3.14 การเลือก NUN ในการทำงานของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1 35

3.15 ค่าความสามารถในการครอบคลุมของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1 36

3.16 ค่าความสามารถในการครอบคลุมของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1 หลังเลือกออบเจ็ค A2 เป็นโปรโตไทป์แล้ว 37

3.17 ค่าความสามารถในการครอบคลุมของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1 หลังเลือกออบเจ็ค A2 และ B3 เป็นโปรโตไทป์แล้ว 37

4.1 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.2 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B1 เป็นโปรโตไทป์แล้ว 61

4.3 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค A3 และ B1 เป็นโปรโตไทป์แล้ว 61

4.4 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 78

4.5 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B2 เป็นโปรโตไทป์แล้ว 79

4.6 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B2 และ A1 เป็นโปรโตไทป์แล้ว 79

4.7 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B2, A1 และ A4 เป็นโปรโตไทป์แล้ว 80

5.1 รายละเอียดชุดข้อมูลขนาดเล็ก 85

5.2 รายละเอียดชุดข้อมูลขนาดกลาง 86

5.3 รายละเอียดชุดข้อมูลขนาดใหญ่ 86

5.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล IRIS 88

5.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล ECOLI 89

5.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล CREDIT 90

5.7 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล YEAST 91

5.8 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล MUSK 92

5.9 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล PEN 93

5.10 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม 94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

5.11 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS 98
5.12 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs 99
5.13 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs 100
5.14 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ 101
ก.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 1 118
ก.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 2 118
ก.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 3 118
ก.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 4 119
ก.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 5 119
ก.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 1 119
ก.7 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 2 120
ก.8 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 3 120
ก.9 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 4 120
ก.10 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 5 121
ข.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 1 123
ข.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 2 123
ข.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 3 123
ข.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 4 124
ข.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 5 124
ข.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 1 124
ข.7 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 2 125
ข.8 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 3 125
ข.9 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 4 125
ข.10 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 5 126
จ.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกจากเซตย่อย ของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อยของทั้ง 4 ชุดข้อมูล 137
จ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกจากเซตย่อย ของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมดของทั้ง 4 ชุดข้อมูล 137

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

ฉ.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” ของ HGAs แบบที่ใช้เทคนิค “Singleton” ของทั้ง 4 ชุดข้อมูล 141

ฉ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” ของ HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ของทั้ง 4 ชุดข้อมูล 141

ช.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” ของ HGAs แบบที่ใช้ทั้งเทคนิค “Diversity Crossover” และเทคนิค “Diversity Selection” ของทั้ง 4 ชุดข้อมูล 145

ช.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” ของ HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Crossover” ของทั้ง 4 ชุดข้อมูล 146

ช.3 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” ของ HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Selection” ของทั้ง 4 ชุดข้อมูล 146

ช.4 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” ของ HGAs แบบที่ไม่ใช้ทั้งเทคนิค “Diversity Crossover” และเทคนิค “Diversity Selection” ของทั้ง 4 ชุดข้อมูล 147

ซ.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ลำดับการทำงานของ HGAs แบบที่ ซ.1 152

ซ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ลำดับการทำงานของ HGAs แบบที่ ซ.2 153

ฌ.1 ค่าตัวแปรของ GAs แบบต่างๆ 157

ฌ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.1 158

ฌ.3 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.2 158

ฌ.4 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.3 159

ฌ.5 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.4 159

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 แผนภาพแสดงการสร้างความความสัมพันธ์ระหว่างออบเจ็คกับชนิดของข้อมูล 9
2.2 แผนภาพแสดงกระบวนตัดสินใจ 10
2.3 ตัวอย่างการจำแนกข้อมูลโดยใช้ NN Rule 11
2.4 ออบเจ็ค x และออบเจ็ค y บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ 13
2.5 ตัวอย่างของข้อมูลที่ไม่เข้าพวก 14
2.6 แผนผังเทคนิคการเลือกโปรโตไทป์ แบบต่างๆ 15
2.7 ขอบเขตการตัดสินใจในรูปของ Voronoi Diagram 16
3.1 แผนภาพขั้นตอนการทำงานของวิธี MCS 19
3.2 ชุดข้อมูลตัวอย่างที่ 1 บนปริภูมิคุณลักษณะเฉพาะ 1 มิติ 20
3.3 การจำแนกชนิดของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 1 โดยใช้ NN Rule 21
3.4 ชุดข้อมูลตัวอย่างที่ 2 บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ 23
3.5 การเรียงลำดับข้อมูลของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 24
3.6 การเลือก NUN ในรอบแรกของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 25
3.7 ออบเจ็คที่สามารถครอบคลุมออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 ได้ 26
3.8 ส่วนหนึ่งของตารางความรอบคลุมในรอบแรกของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 26
3.9 การเลือก NUN ในรอบที่สองของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 31
3.10 แผนภาพวัฏจักรการทำงานของ GAs 39
3.11 ตัวอย่างการเข้ารหัสโครโมโซม 40
3.12 แผนภาพขั้นตอนการทำงานของ GAs 43
3.13 การสร้างประชากรเริ่มต้น 45
3.14 ตัวอย่างการสุ่มกำหนดค่าเริ่มต้นให้กับแต่ละบิตของโครโมโซมในประชากร 45
3.15 การกำหนดค่าเริ่มต้นให้โครโมโซมที่ดีที่สุด 46
3.16 ตัวอย่างขั้นตอนการสร้างรูเล็ทวีล 48
3.17 แผนภูมิวงกลมของตัวอย่างการสร้างรูเล็ทวีล 48
3.18 ตัวอย่างการทำงานของกระบวนการครอสโอเวอร์ของ GAs 49
3.19 ตัวอย่างการทำงานของกระบวนการกลายพันธุ์ของ GAs 50
3.20 ตัวอย่างการสร้างกลุ่มโครโมโซมร่วมของ GAs 50
3.21 ตัวอย่างการทำงานของกระบวนการเลือกสรรประชากรในรุ่นถัดไปของ GAs 51

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.1 แผนภาพการทำงานของวิธี MCS และ GAs 54
4.2 แผนภาพการทำงานของ HGAs 55
4.3 แผนภาพรายละเอียดการทำงานของ HGAs 56
4.4 ตัวอย่างประชากรเริ่มต้นที่ได้จากกระบวนการกำหนดค่าเริ่มต้นของ HGAs 58
4.5 ตัวอย่างการทำงานของกระบวนการ MCS ของ HGAs 62
4.6 ตัวอย่างโปรโตไทป์ผลลัพธ์ที่ได้จากกระบวนการ MCS 62
4.7 ตัวอย่างการทำงานของกระบวนการครอสโอเวอร์ของ HGAs 63
4.8 ตัวอย่างการทำงานของกระบวนการกลายพันธุ์ของ HGAs 64
4.9 ตัวอย่างการทำงานของกระบวนการเลือกสรรประชากรในรุ่นถัดไปของ HGAs 65
4.10 ส่วนหนึ่งของตารางความรอบคลุมของออบเจ็ค A4 จากชุดข้อมูลตัวอย่างที่ 2 66
4.11 แผนผังการทำงานของ HGAs ที่ใช้เทคนิค "Singleton" 66
4.12 ประชากรตัวอย่างที่ 1 68
4.13 ประชากรตัวอย่างที่ 2 69
4.14 ความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2 70
4.15 ความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 3 70
4.16 ความหลากหลายระหว่าง CHROMOSOME 2 กับ CHROMOSOME 3 70
4.17 ตัวอย่างโครโมโซมที่สุ่มเลือกได้ 71
4.18 ตัวอย่างโครโมโซมที่สุ่มเลือกใหม่ 72
4.19 ตัวอย่างโครโมโซมในประชากร และโครโมโซมลูก 73
4.20 ตัวอย่างการสร้างกลุ่มโครโมโซมร่วมของ HGAs 74
4.21 ตัวอย่างการเลือกโครโมโซมโดยใช้ค่าความเหมาะสม 74
4.22 ตัวอย่างการเลือกโครโมโซมโดยใช้ค่าความหลากหลาย 76
5.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม 95
5.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ 102
5.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ ของทุกชุดข้อมูล 103
5.4 แผนภูมิแสดงลำดับเครือญาติของตัวอย่างการทดลองที่ 1 105
5.5 ขอบเขตระหว่างเจเนอเรชั่นที่ 1 กับเจเนอเรชั่นที่ 2 106
5.6 โครโมโซมลำดับที่ 1 ในเจเนอเรชั่นที่ 1 (มีค่า $Size(S)$ =16 และค่า $Accuracy(S)$ =150) 106

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

5.7 กระบวนการเลือกประชากรในรุ่นถัดไปทำการเลือกโครโมโซมนี้ด้วยค่า “Fitness Value” เป็นลำดับที่ 3 เพื่อนำไปเป็นโครโมโซมลำดับที่ 3 ในเจเนอเรชั่นถัดไป 107

5.8 กระบวนการเลือกประชากรในรุ่นถัดไปทำการเลือกโครโมโซมนี้ด้วยค่า “Diversity Value” เป็นลำดับที่ 4 เพื่อนำไปเป็นโครโมโซมลำดับที่ 9 ในเจเนอเรชั่นถัดไป 107

5.9 การเกิดกระบวนการทางพันธุกรรมระหว่างโครโมโซมลำดับที่ 7 กับโครโมโซลำดับที่ 5 เป็นผลให้ได้โครโมโซมลูกลำดับที่ 14 108

5.10 ลักษณะการพัฒนาของโครโมโซม 108

ค.1 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล IRIS 127

ค.2 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล ECOLI 128

ค.3 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล CREDIT 129

ค.4 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล YEAST 130

ค.5 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล MUSK 131

ค.6 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล PEN 132

ง.1 โครงสร้างครอบครัวของตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS 134

จ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกจากเซตย่อย (เฉพาะกรณีสอดคล้อง 100%) 138

ฉ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของเทคนิค “Singleton” (เฉพาะกรณีสอดคล้อง 100%) 142

ช.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของเทคนิค “Diversity” (เฉพาะกรณีสอดคล้อง 100%) 148

ซ.1 แผนภาพการทำงานของ HGAs แบบที่ ซ.1 150

ซ.2 แผนภาพการทำงานของ HGAs แบบที่ ซ.2 151

ซ.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของลำดับการทำงาน (เฉพาะกรณีสอดคล้อง 100%) .. 154

ฌ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าผลลัพธ์ตัวแปรของ GAs (เฉพาะกรณีสอดคล้อง 100%) 160

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อย่างไรก็ตามจากข้อดีของการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule ที่ไม่ต้องทำการสร้างแบบจำลองก็ทำให้มีข้อด้อยตามมา เช่น การใช้การจำแนกข้อมูลวิธีนี้ไปใช้จำแนกข้อมูลที่มีจำนวนคุณลักษณะเฉพาะมาก ก็จำเป็นต้องใช้ Training Dataset ที่มีข้อมูลจำนวนมากตามไปด้วย (เรียกปัญหาลักษณะนี้ว่า "Curse of Dimensionality" [1][3]) ปัญหานี้สามารถแก้ไขได้โดยการลดจำนวนมิติลง (Dimensionality Reduction) เช่น วิธี Feature Exaction หรือ Feature Selection [1][3] เป็นต้น หรือจะใช้วิธีที่ง่ายกว่านั้นก็คือการสร้างข้อมูลที่ใช้อ้างอิงขึ้นมาใหม่จนมีจำนวนมากพอกับความต้องการ แต่ก็จะเสียเวลาในการคำนวณเพื่อจำแนกข้อมูลมาก (High Computational Demand) อีกทั้งยังต้องใช้หน่วยความจำจำนวนมาก (High Memory Demand) เพื่อใช้ในการเก็บข้อมูลเหล่านั้น

กระบวนการเลือกโปรโตไทป์เป็นวิธีการหนึ่งที่ถูกนำมาใช้เพื่อลดปัญหาดังกล่าว โดยจะทำการลดจำนวนข้อมูลใน Training Dataset ลงให้มากที่สุดเท่าที่จะทำได้ ในขณะที่ยังคงรักษาประสิทธิภาพในการจำแนกให้ใกล้เคียงกับการใช้ข้อมูลใน Training Dataset เดิมเป็นข้อมูลอ้างอิง

วิทยานิพนธ์ฉบับนี้จึงได้นำเสนอ "การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม"

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

วิทยานิพนธ์ฉบับนี้มุ่งหวังเพื่อศึกษาการลดจำนวนข้อมูลใน Training Dataset ลงให้มากที่สุดเท่าที่จะทำได้ ในขณะที่ยังคงรักษาประสิทธิภาพในการจำแนกให้ใกล้เคียงกับการใช้ข้อมูลใน Training Dataset เดิมเป็นข้อมูลอ้างอิง เพื่อลดปัญหาความจำเป็นต้องใช้การคำนวณในปริมาณสูงและความต้องการหน่วยความจำในการใช้งานเป็นจำนวนมากที่เกิดขึ้นในการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule ในกรณีที่ Training Dataset มีข้อมูลจำนวนมาก

1.3 สมมติฐานของการศึกษา

เนื่องจากการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule จำเป็นที่จะต้องเก็บข้อมูลใน Training Dataset ไว้ทั้งหมด ดังนั้นในกรณีที่ Training Dataset มีข้อมูลจำนวนมาก ก็จะต้องเสียเวลาในการคำนวณมาก อีกทั้งยังต้องใช้หน่วยความจำจำนวนมากเพื่อใช้ในการเก็บข้อมูลเหล่านั้น

ดังนั้นจึงตั้งสมมุติฐานว่าถ้าสามารถใช้เพียงข้อมูลบางตัว แล้วทดแทนข้อมูลทั้งหมดใน Training Dataset ได้ (โดยการทดแทนกันได้คือจะต้องมีประสิทธิภาพการจำแนกใกล้เคียงกัน) ก็จะสามารถช่วยลดปัญหาที่เกิดขึ้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

แนวคิดหลักที่ใช้ในวิทยานิพนธ์ฉบับนี้ คือ แนวคิด "ความครอบคลุม" ที่ได้นำเสนอไว้ในบทความที่มีชื่อว่า "Minimal Consistent Set (MCS) Identification for Optimal Nearest Neighbor - Decision Systems Design" [4] โดย Belur V. Dasarathy

จากสมมติฐานที่ว่าถ้าสามารถใช้เพียงข้อมูลบางตัวที่สามารถทดแทนข้อมูลทั้งหมดใน Training Dataset ได้ Belur ได้นำเสนอว่าจริงๆ แล้วข้อมูลใน Training Dataset นั้นมีบางตัวที่ซ้ำซ้อนกันอยู่ ทำให้จำนวนข้อมูลมีมากกว่าที่ควรจะเป็น ดังนั้น Belur จึงเสนอหลักการ "ความครอบคลุม" โดยแนวคิดความครอบคลุมนี้จะทำให้ทราบความสัมพันธ์ของข้อมูลแต่ละตัวใน Training Dataset ซึ่งจะทำให้สามารถเลือกเฉพาะข้อมูลบางตัวที่มีความสำคัญ และสามารถลดข้อมูลที่ซ้ำซ้อนลงได้

อีกหลักการหนึ่งที่วิทยานิพนธ์ฉบับนี้นำมาใช้ คือ Genetic Algorithms (GAs) ซึ่งเป็นวิธีการทางวิทยาศาสตร์ที่ใช้ในการหาผลลัพธ์โดยการประมาณค่าเพื่อใช้ในการหาคำตอบที่ดีที่สุด (Optimization) ที่สามารถนำไปประยุกต์ใช้ในการแก้ปัญหาต่างๆ ได้มากมาย โดยการนำไปแก้ปัญหาแต่ละแบบก็จะต้องประยุกต์ใช้ GAs ในรูปแบบที่แตกต่างกันออกไป โดยในวิทยานิพนธ์ฉบับนี้จะนำ GAs มาประยุกต์ใช้ในการเลือกโปรโตไทป์ โดยอ้างอิงจากบทความที่มีชื่อว่า "Nearest Prototype Classification: Clustering, Genetic Algorithms, or Random Search?" [5] โดย Ludmila I. uncheva และ James C. Bezdek

1.5 ขอบเขตการวิจัย

วิทยานิพนธ์ฉบับนี้ทำการศึกษา และพัฒนาวิธีการเลือกโปรโตไทป์ โดยกำหนดขอบเขตของการวิจัย ดังนี้

1. การเลือกโปรโตไทป์ในวิทยานิพนธ์ฉบับนี้เป็นการเลือกโปรโตไทป์เพื่อลดจำนวนข้อมูลใน Training Dataset ลงให้มากที่สุดเท่าที่จะทำได้ โดยที่ยังคงรักษาประสิทธิภาพในการจำแนกใกล้เคียงเดิม และผลลัพธ์ที่ได้จะเป็นเซตย่อยของ Training Dataset

2. การวัดความใกล้เคียงของประสิทธิภาพการจำแนกในวิทยานิพนธ์นี้จะยึด "คุณสมบัติความสอดคล้อง" เป็นหลัก

3. ชุดข้อมูลที่ใช้ในวิทยานิพนธ์ฉบับนี้เป็นชุดข้อมูลมาตรฐานจาก UCI Site [6] โดยเลือกใช้ชุดข้อมูลที่มีค่าคุณลักษณะเฉพาะที่เป็นจำนวนจริง และมีค่าคุณลักษณะครบทุกถ้วนข้อมูลเท่านั้น โดยถือว่าเป็นชุดข้อมูลที่มีความถูกต้อง และไม่มีข้อมูลที่ผิดพลาด (Noise)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6 ขั้นตอนการศึกษา

1. ศึกษาทฤษฎี และความรู้พื้นฐานที่เกี่ยวข้องกับการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule

2. ศึกษาทฤษฎี และความรู้พื้นฐานที่เกี่ยวข้องกับการเลือกโปรโตไทป์

3. ศึกษาการเลือกโปรโตไทป์ของงานวิจัยที่เกี่ยวข้องที่มีผู้นำเสนอมาก่อนหน้านี้

4. ศึกษาทฤษฎี และความรู้พื้นฐานที่เกี่ยวข้องกับวิธี MCS

5. ศึกษาทฤษฎี และความรู้พื้นฐานที่เกี่ยวข้องกับ GAs

6. ศึกษาการนำ GAs มาประยุกต์ใช้ในการเลือกโปรโตไทป์ของงานวิจัยที่เกี่ยวข้องที่ มีผู้นำเสนอมาก่อนหน้านี้

7. ทดลองเพื่อเปรียบเทียบผลลัพธ์จากการเลือกโปรโตไทป์ เพื่อศึกษาข้อดี และข้อด้อยของวิธีต่างๆ

8. พัฒนาการเลือกโปรโตไทป์โดยใช้วิธี MCS และ GAs

9. สรุปผลการทำงานของวิธีที่พัฒนาขึ้น พร้อมวิเคราะห์ผลที่ได้

10. จัดทำเอกสารประกอบวิทยานิพนธ์

1.7 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย

1. เครื่องคอมพิวเตอร์ส่วนบุคคล
   - หน่วยประมวลผลกลาง ยี่ห้อ Intel รุ่น Core2Duo ความเร็ว 3.0 GHz
   - หน่วยความจำหลักขนาด 2 GB
2. ระบบปฏิบัติการ Windows XP Professional (SP2)
3. โปรแกรม Microsoft Visual C++ 6.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.8 โครงสร้างของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 6 บท ดังนี้

บทที่ 1 กล่าวถึงความเป็นมาของงานวิจัย ความมุ่งหมาย วัตถุประสงค์ สมมติฐาน ทฤษฎีที่ใช้ ขอบเขตของการวิจัย ขั้นตอนการศึกษา ตลอดจนเครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย

บทที่ 2 กล่าวถึงนิยามและทฤษฎีพื้นฐานที่เกี่ยวข้องกับการเลือกโปรโตไทป์ ทั้งสัญลักษณ์และสมการในคณิตศาสตร์ต่างๆ ที่จะถูกใช้ในการอธิบายในหัวข้อถัดไป

บทที่ 3 กล่าวถึงงานวิจัยที่เกี่ยวข้อง

บทที่ 4 กล่าวถึงแนวคิด และหลักการทำงานของการเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม

บทที่ 5 กล่าวถึงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธีการที่นำเสนอนี้กับวิธี MCS และ GAs

บทที่ 6 กล่าวถึงบทสรุปผลการวิจัย และข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
นิยามและทฤษฎีพื้นฐานที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงนิยามและทฤษฎีพื้นฐานที่เกี่ยวข้องกับการเลือกโปรโตไทป์ ทั้งสัญลักษณ์ และสมการในคณิตศาสตร์ต่างๆ ที่จะถูกใช้ในการอธิบายในหัวข้อถัดไป

2.1 นิยามพื้นฐาน (Basic Definition)

หัวข้อนี้จะแสดงนิยามพื้นฐานที่ควรทราบในการศึกษาการเลือกโปรโตไทป์ ที่จะถูกใช้ในการอธิบายต่อไป โดยมีรายละเอียดดังนี้

2.1.1 ออบเจ็ค (Object)

ออบเจ็ค คือ ข้อมูลที่อยู่ใน Training Dataset หรือ Testing Dataset ซึ่งอาจจะเป็นคน สัตว์ หรือวัตถุใดๆ ที่กำลังสนใจ ซึ่งข้อมูลที่แสดงอยู่ในโลกแห่งความจริงนั้นไม่สามารถนำมาใช้คำนวณได้โดยตรง แต่จะต้องแทนวัตถุเหล่านั้นด้วยค่าที่วัดได้จากกคุณลักษณะเฉพาะ (Feature Value) ของวัตถุนั้นๆ

ตัวอย่างเช่น สมมุติว่ากำลังสนใจออบเจ็คที่เป็นมนุษย์ ก็อาจจะพิจารณาด้วยค่าคุณลักษณะเฉพาะเพียง 2 ค่า คือ ความสูง (ซม.) และน้ำหนัก (กก.) ซึ่งจะทำให้สามารถพิจารณามนุษย์เป็นเพียงข้อมูลที่มี 2 มิติ นั่นคือ ความสูง และน้ำหนัก และเมื่อทำการแทนวัตถุใดๆ ด้วยข้อมูลที่มีจำนวนมิติเท่าจำนวนคุณลักษณะเฉพาะของวัตถุนั้นๆ แล้ว ก็จะเสมือนการพิจารณาวัตถุเหล่านั้นบนจุดพิกัดระบุตำแหน่ง (Coordinate) บนปริภูมิคุณที่มีมิติขนาด d (R^d) ที่ครอบคลุมจุดพิกัดของออบเจ็คทั้งหมดนั่นเอง โดยถ้ากำหนดให้ออบเจ็คใดๆ แทนด้วย x ก็จะสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Object \ : \ x \in \Re^d \tag{2.1}$$

การเลือกคุณลักษณะเฉพาะที่ใช้แสดงถึงวัตถุจะแตกต่างกันไปขึ้นกับการใช้งาน เช่น การเลือกแค่ความสูงกับน้ำหนักอาจจะเพียงพอต่อการนำไปใช้จำแนกว่ามนุษย์คนนั้นมีรูปร่างสมส่วนหรือไม่ แต่ถ้าต้องการที่จะจำแนกว่ามนุษย์คนนั้นเป็นคนเชื้อสายใด อาจจะต้องใช้คุณลักษณะเฉพาะอย่างอื่นเพิ่มเข้ามาร่วมด้วย เช่น สีผิว สีผม หรือสีนัยน์ตา เป็นต้น ซึ่งจะเห็นว่าวัตถุก็คือมนุษย์เหมือนกันแต่เป้าหมายการใช้งานแตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 ปริภูมิคุณลักษณะเฉพาะ (Feature Space)

ปริภูมิคุณลักษณะเฉพาะเป็นปริภูมิที่ครอบคลุมจุดพิกัดของออบเจ็คทั้งหมดใน Training Dataset และมีมิติเท่ากับจำนวนคุณลักษณะเฉพาะที่ใช้ในการพิจารณาออบเจ็คเหล่านั้น โดยออบเจ็คบนปริภูมิคุณลักษณะเฉพาะจะอยู่ในรูปของคู่อันดับออบเจ็ค-คลาส (Object-Class Pairs) โดยคู่อันดับตัวหน้าจะเป็นค่าคุณลักษณะเฉพาะ หรือพิกัดของออบเจ็คที่มีจำนวนมิติเท่ากับจำนวนคุณลักษณะ ส่วนคู่อันดับตัวหลังจะเป็นชนิดของออบเจ็ค โดยถ้ากำหนดให้ออบเจ็คใดๆ แทนด้วย x จำนวน n ตัว ก็จะสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Object = \{((x_{11}, x_{12}, ..., x_{1d}), \theta_1), ((x_{21}, x_{22}, ..., x_{2d}), \theta_2), ..., ((x_{n1}, x_{n2}, ..., x_{nd}), \theta_n)\} \quad (2.2)$$

2.1.3 ปริภูมิคำตอบ (Solution Space)

ปริภูมิคำตอบเป็นปริภูมิของคำตอบทุกคำตอบที่เป็นไปได้ในการเลือกโปรโตไทป์ที่ให้ผลลัพธ์เป็นเซตย่อย (ซึ่งจะอธิบายต่อไปในหัวข้อที่ 2.5.4) กล่าวคือทำการเลือกโปรโตไทป์จากออบเจ็คทั้งหมดใน Training Dataset ดังนั้นขนาดของปริภูมิคำตอบก็คือวิธีการเลือกที่เป็นไปได้ทั้งหมด ซึ่งจะแปรผันตรงกับจำนวนออบเจ็คใน Training Dataset เช่น Training Dataset มีข้อมูลทั้งหมดจำนวน 10 ออบเจ็ค (การเลือกโปรโตไทป์ของแต่ละออบเจ็คก็จะเป็นไปได้ 2 กรณีก็คือเลือกหรือไม่เลือกเป็นโปรโตไทป์) ดังนั้นวิธีการเลือกโปรโตไทป์ที่เป็นไปได้ทั้งหมดเท่ากับ 2^{10} หรือ 1024 วิธี

2.1.4 ชุดข้อมูลฝึกสอน (Training Dataset)

ชุดข้อมูลฝึกสอนเป็นชุดข้อมูลที่ใช้ในการอ้างอิงในการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule โดยข้อมูลใน Training Dataset จะต้องทราบค่าคุณลักษณะเฉพาะ และชนิดของแต่ละข้อมูล ซึ่งใช้เป็นอินพุตในการเลือกโปรโตไทป์ โดยถ้ากำหนดให้ Training Dataset มีสมาชิก คือ *traindata* จำนวน n ตัว ก็จะสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Training\ DataSet = \{traindata_1, traindata_2, ..., traindata_n\} \quad (2.3)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.5 ชุดข้อมูลทดสอบ (Testing Dataset)

ชุดข้อมูลทดสอบเป็นชุดข้อมูลที่ใช้ในการทดสอบการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule โดยข้อมูลใน Testing Data จะทราบแต่ค่าคุณลักษณะเฉพาะแต่จะไม่ทราบชนิดของข้อมูล โดยถ้ากำหนดให้ Testing Dataset มีสมาชิก คือ *testdata* จำนวน n ตัว ก็จะสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Testing\ DataSet = \{testdata_1, testdata_2, ..., testdata_n\} \tag{2.4}$$

2.1.6 เซตชนิดของข้อมูล (Category Set)

เซตชนิดของข้อมูลเป็นเซตของชื่อที่เป็นตัวระบุชนิดที่เป็นไปได้ทั้งหมดของออบเจ็ค โดยถ้ากำหนดให้เซตชนิดของข้อมูลมีสมาชิก คือ Class ที่แทนด้วยสัญลักษณ์ θ จำนวน c ตัว ก็จะสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Category\ Set = \{\theta_1, \theta_2, ..., \theta_c\} \tag{2.5}$$

2.1.7 โปรโตไทป์ผลลัพธ์ (Prototype Set)

โปรโตไทป์ผลลัพธ์เป็นผลลัพธ์ที่ได้จากกระบวนการเลือกโปรโตไทป์ ซึ่งเป็นออบเจ็คที่ใช้เป็นข้อมูลอ้างอิงในการจำแนกข้อมูลโดยใช้ Nearest Neighbor Rule โดยออบเจ็คเหล่านี้จะต้องทราบค่าคุณลักษณะเฉพาะ และทราบชนิดของแต่ละออบเจ็คซึ่งจะอยู่ในรูปของคู่อันดับออบเจ็คคลาส โดยถ้ากำหนดให้โปรโตไทป์ใดๆ แทนด้วย x และชนิดที่แทนด้วยสัญลักษณ์ θ จำนวน n ตัว ก็จะสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Prototype = \{(x_1, \theta_1), (x_2, \theta_2), ..., (x_n, \theta_n)\} \tag{2.6}$$

2.1.8 ออบเจ็คไม่ทราบชนิด (Unknown Object)

ออบเจ็คไม่ทราบชนิดเป็นออบเจ็คที่ทราบค่าคุณลักษณะเฉพาะแต่ไม่ทราบชนิดของออบเจ็ค โดยถ้ากำหนดให้โปรโตไทป์ใดๆ แทนด้วย x และชนิดที่แทนด้วยสัญลักษณ์ θ ก็จะสามารถเขียนอยู่ในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Unknown\ Object = (x, \theta) \tag{2.7}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.9 ฟังก์ชั่นคำนวณระยะทาง (Distance Function)

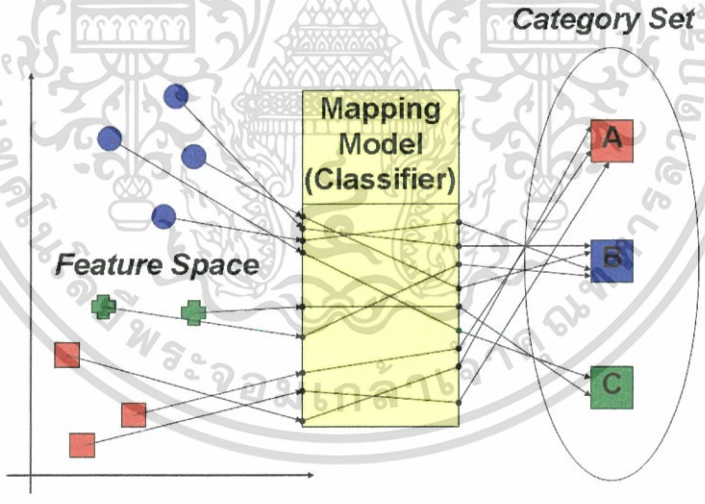
ฟังก์ชั่นคำนวณระยะทางเป็นฟังก์ชั่นที่ใช้วัดระยะห่างระหว่างออบเจ็ค (จุดพิกัด) 2 ออบเจ็คที่อยู่ภายบนปริภูมิคุณลักษณะเฉพาะเดียวกัน โดยคำนวณจากค่าคุณลักษณะเฉพาะของทั้ง 2 ออบเจ็ค โดยถ้าออบเจ็คใดๆ มีระยะห่างกันมาก ก็จะหมายความว่ามีความแตกต่างกันมาก ในทางกลับกันถ้าออบเจ็คใดๆ มีระยะห่างกันน้อย ก็จะหมายความว่ามีความแตกต่างกันน้อย

โดยจะกำหนดให้แสดงฟังก์ชั่นวัดระยะทางด้วยสัญลักษณ์ d และกำหนดให้ออบเจ็คใดๆ แทนด้วย x ซึ่งสามารถเขียนอยู่ในรูปสมการคณิตศาสตร์ได้ดังนี้

$$Distance\ Function\ :\ d(x_i, x_j) \mid x_i, x_j \in \Re^d \tag{2.8}$$

2.2 การจำแนกข้อมูล (Classification)

การจำแนกข้อมูลสามารถพิจารณาได้ 2 แบบดังนี้ แบบแรกเป็นการพิจารณาเสมือนเป็นการสร้างความสัมพันธ์เพื่อเชื่อมโยง (Mapping Model) จาก “ออบเจ็ค” ที่อยู่ใน “ปริภูมิคุณลักษณะเฉพาะ” ไปยัง “ชนิด” ที่อยู่ใน “เซตชนิดของข้อมูล” ที่เป็นตัวระบุชนิดของออบเจ็คทั้งหมด [1]

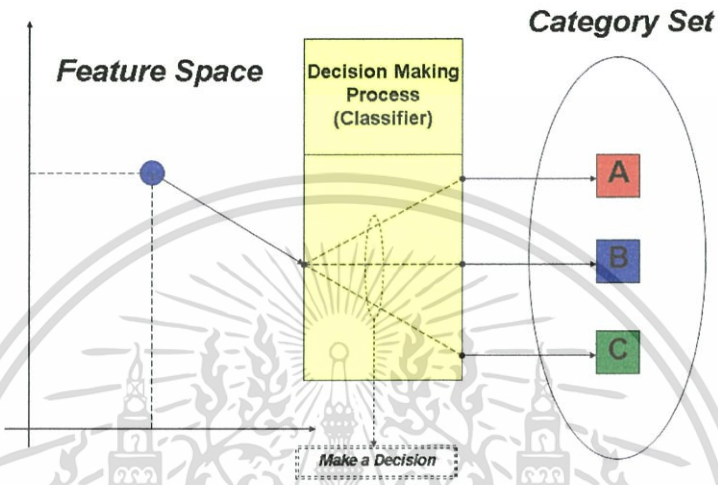


รูปที่ 2.1 แผนภาพแสดงการสร้างความความสัมพันธ์ระหว่างออบเจ็คกับชนิดของข้อมูล

รูปที่ 2.1 แสดงแผนภาพแสดงการสร้างความความสัมพันธ์ระหว่างออบเจ็คกับชนิดของข้อมูลระหว่างออบเจ็คบนปริภูมิคุณลักษณะเฉพาะกับชนิดในเซตชนิดของข้อมูล โดยออบเจ็คจะอยู่บนปริภูมิคุณลักษณะเฉพาะ 2 มิติทางด้านซ้ายมือ จากนั้นตัวจำแนกข้อมูล (Classifier) จะทำการเชื่อมโยงเพื่อบอกความสัมพันธ์ระหว่างออบเจ็คบนปริภูมิคุณลักษณะเฉพาะ 2 มิติทางด้านซ้ายมือไปยังชนิดภายในเซตชนิดของข้อมูลที่แทนด้วยสี่เหลี่ยมในวงรีทางด้านขวา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจำแนกข้อมูลยังสามารถมองได้ในอีกแบบหนึ่งเป็นการพิจารณาเสมือนเป็นกระบวนการตัดสินใจ (Decision Making Process) ที่รับค่า “คุณลักษณะเฉพาะ” ของแต่ละออบเจ็คมาเป็นอินพุต แล้วนำค่าคุณลักษณะเฉพาะเหล่านั้นมาใช้ในการตัดสินใจว่าแต่ละออบเจ็คจะเป็นชนิดใด โดยพิจารณาจากชนิดทั้งหมดที่มีอยู่ในเซตชนิดของข้อมูล และให้เอาพุตออกมาเป็น “ชนิด” ของออบเจ็คนั้นๆ [1]



รูปที่ 2.2 แผนภาพแสดงกระบวนตัดสินใจ

จากรูปที่ 2.2 แสดงแผนภาพแสดงกระบวนตัดสินใจ โดยออบเจ็คจะอยู่บนปริภูมิคุณลักษณะเฉพาะ 2 มิติทางด้านซ้ายมือ จากนั้นตัวจำแนกข้อมูลจะทำการตัดสินใจว่าออบเจ็คที่อยู่บนปริภูมิคุณลักษณะเฉพาะ 2 มิติทางด้านซ้ายมือเป็นชนิดใดภายในเซตชนิดของข้อมูลที่แทนด้วยสี่เหลี่ยมในวงรีทางด้านขวา

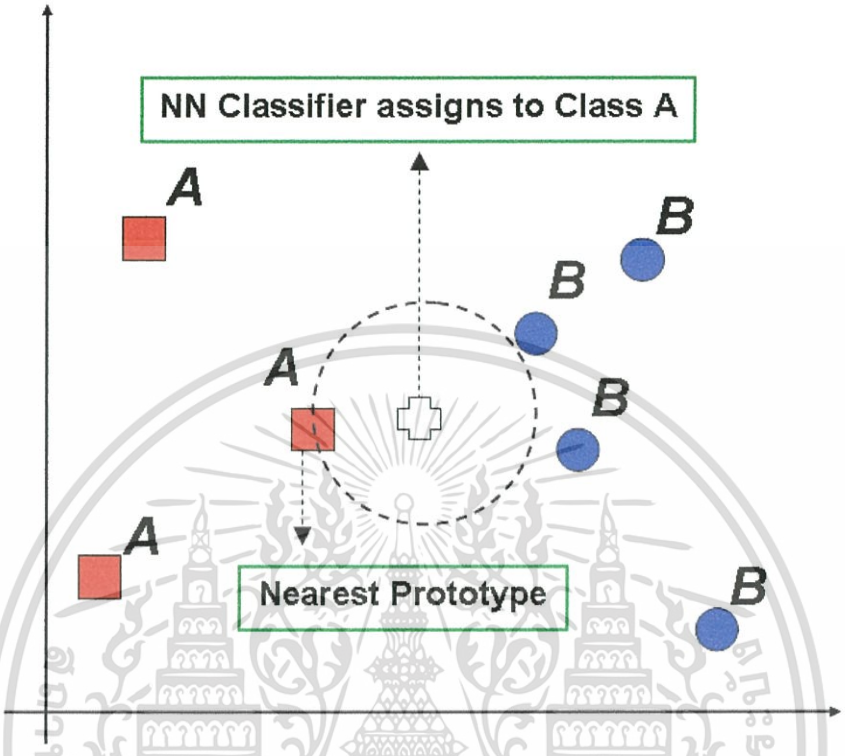
2.3 Nearest Neighbor Rule (NN Rule)

NN Rule เป็นกฎที่ใช้ในการตัดสินใจของการจำแนกข้อมูล โดยเป็นการจำแนกข้อมูลที่ใช้ระยะทางในการตัดสินใจ ข้อดีที่สำคัญของวิธีนี้ คือ ไม่ต้องทำการสร้างแบบจำลองให้ยุ่งยาก ทำให้มีความยืดหยุ่นสูง โดยหลักการทำงานสามารถอธิบายได้ง่ายๆ คือ เมื่อต้องการทราบชนิดของออบเจ็คไม่ทราบชนิด ก็จะทำการเปรียบเทียบว่าออบเจ็คที่ไม่ทราบชนิดมีความเหมือนกับโปรโตไทป์ตัวใดมากที่สุด จากนั้นก็จะทำการกำหนดชนิดของออบเจ็คให้เป็นชนิดเดียวกันกับโปรโตไทป์ที่เหมือนกับออบเจ็คไม่ทราบชนิดนั้นมากที่สุด โดยในการวัดความเหมือนจะใช้ฟังก์ชั่นระยะทางวัดออกมาในรูประยะทาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นิยาม Nearest Neighbor Rule

"กำหนดชนิดให้ออบเจ็คไม่ทราบชนิด เป็นชนิดเดียวกันกับชนิดของโปรโตไทป์ที่อยู่ใกล้กับออบเจ็คไม่ทราบชนิดนั้นมากที่สุด" [1-3]



รูปที่ 2.3 ตัวอย่างการจำแนกข้อมูลโดยใช้ NN Rule

รูปที่ 2.3 แสดงตัวอย่างการจำแนกข้อมูลโดยใช้ NN Rule โดยกำหนดให้ชุดข้อมูลฝึกสอนมีโปรโตไทป์จำนวน 7 ตัว ประกอบไปด้วยโปรโตไทป์ชนิด A จำนวน 3 ตัว (รูปสี่เหลี่ยม) และโปรโตไทป์ชนิด B จำนวน 4 ตัว (รูปวงกลม) ซึ่งโปรโตไทป์ทั้ง 7 ตัว อยู่ในตำแหน่งที่แตกต่างกันบนปริภูมิคุณลักษณะเฉพาะ 2 มิติ แสดงได้ดังรูปที่ 2.3 จากนั้นกำหนดให้มีออบเจ็คที่ไม่ทราบชนิดออบเจ็คหนึ่ง (รูปเครื่องหมายบวก) มีตำแหน่งบนปริภูมิคุณลักษณะเฉพาะเดียวกัน ดังที่แสดงในรูปที่ 2.3 เช่นกัน และเมื่อต้องการจำแนกประเภทของออบเจ็คไม่ทราบชนิดว่าเป็นข้อมูลชนิดใดโดยใช้ NN Rule ก็จะทำการตัดสินใจกำหนดให้ออบเจ็คไม่ทราบชนิดนี้เป็นชนิดเดียวกันกับชนิดของโปรโตไทป์ที่อยู่ใกล้กับออบเจ็คนั้นๆ มากที่สุด ซึ่งจากรูปจะเห็นว่าการจำแนกข้อมูลโดยใช้ NN Rule จะทำการกำหนดให้ออบเจ็คไม่ทราบชนิดเป็นชนิด A เช่นเดียวกับชนิดของโปรโตไทป์ที่อยู่ใกล้กับออบเจ็คนั้นๆ มากที่สุด ซึ่งสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

$$\theta = \theta' \mid (x',\theta') \in \textit{Prototype Set}\ (x',y) = \min_{i=1}^{n} d(x_i,y) \qquad (2.9)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยกำหนดให้

- x คือ โปรโตไทป์จำนวน n ตัว
- y คือ ออบเจ็คไม่ทราบชนิด
- x' คือ โปรโตไทป์ที่อยู่ใกล้กับ y มากที่สุดและมีชนิดเป็น θ'

2.4 การวัดระยะทางแบบยูคลิด (Euclidean Distance)

การวัดระยะทางแบบยูคลิด [1] เป็นการคำนวณระยะห่างระหว่าง 2 วัตถุใดๆ ซึ่งในวิทยานิพนธ์ฉบับนี้จะใช้การวัดระยะทางแบบยูคลิดเป็นเครื่องมือในการวัดความเหมือนระหว่างออบเจ็ค โดยยิ่งระยะห่างระหว่างออบเจ็คน้อยก็หมายความว่ามีความแตกต่างกันน้อย (หรือพูดอีกนัยหนึ่งว่ามีความเหมือนกันมาก) ในทางกลับกันยิ่งระยะห่างระหว่างออบเจ็คมากก็หมายความว่ามีความแตกต่างกันมาก (หรือพูดอีกนัยหนึ่งว่ามีความเหมือนกันน้อย) ซึ่งสามารถแสดงในรูปสมการคณิตศาสตร์ได้ดังนี้

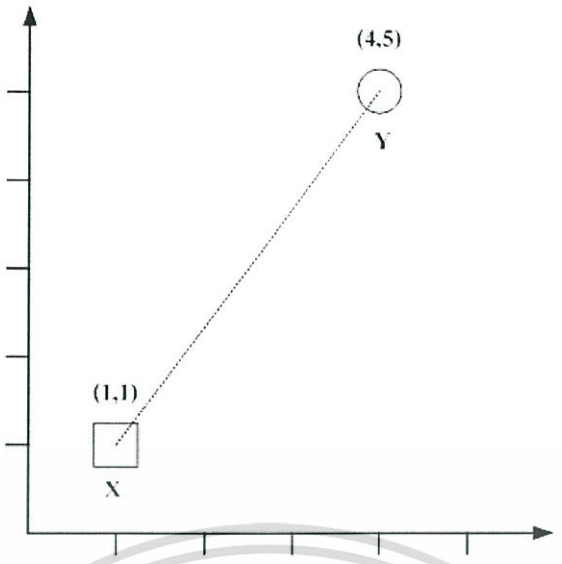
$$d(x,y)=\sqrt{\sum_{k=1}^{n}(x_k-y_k)^2} \tag{2.10}$$

โดยกำหนดให้

- $d(x,y)$ คือ ระยะห่างระหว่างออบเจ็คบนปริภูมิคุณลักษณะเฉพาะ
- x_k และ y_k คือ ค่าคุณลักษณะเฉพาะของออบเจ็ค x และออบเจ็ค y ตามลำดับ
- k คือ จำนวนค่าลักษณะเฉพาะที่ไปได้ทั้งหมด แทนด้วยจำนวนจริงขนาด n มิติ

การวัดระยะทางแบบยูคลิดเหมาะในการคำนวณระยะห่างระหว่างออบเจ็คที่มีค่าคุณลักษณะเฉพาะที่เป็นจำนวนจริง และมีจำนวนของค่าคุณลักษณะเฉพาะที่ไม่แตกต่างกันมากนัก ตัวอย่างเช่น กำหนดให้ออบเจ็ค x มีคุณลักษณะเฉพาะจำนวน 2 มิติมีค่าเท่ากับ (1, 1) และออบเจ็ค y มีคุณลักษณะเฉพาะจำนวน 2 มิติมีค่าเท่ากับ (4, 5) ซึ่งสามารถแสดงได้ดังรูปที่ 2.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.4 ออบเจ็ค x และออบเจ็ค y บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ

จากรูปที่ 2.4 แสดงออบเจ็ค x และออบเจ็ค y บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ โดยออบเจ็ค x แทนด้วยรูปสี่เหลี่ยม และออบเจ็ค y แทนด้วยรูปวงกลม ดังนั้นสามารถคำนวณการวัดระยะทางแบบยูคลิดระหว่างออบเจ็ค x กับออบเจ็ค y ได้ดังนี้

$$
\begin{aligned}
&= \sqrt{(1-4)^2 + (1-5)^2} \\
&= \sqrt{(-3)^2 + (-4)^2} \\
&= \sqrt{9+16} \\
&= \sqrt{25} \\
&= 5 \text{ หน่วย}
\end{aligned}
$$

2.5 การเลือกโปรโตไทป์ (Prototype Selection)

จากสมการที่ (2.9) จะเห็นว่าการจำแนกข้อมูลโดยใช้ NN Rule นั้นต้องทำการคำนวณระยะห่างระหว่างออบเจ็คไม่ทราบชนิดกับโปรโตไทป์ทั้งหมด ดังนั้นในกรณีที่โปรโตไทป์มีจำนวนมาก ก็จะเสียเวลาในการคำนวณเพื่อจำแนกข้อมูลมาก อีกทั้งยังต้องการหน่วยความจำในการใช้งานเป็นจำนวนมากเพื่อที่จะให้เพียงพอในการเก็บข้อมูลเหล่านั้น จึงต้องมีการเลือกโปรโตไทป์เพื่อลดจำนวนข้อมูลเหล่านั้นลง

นอกจากนี้แล้วการเลือกโปรโตไทป์ไม่ได้ใช้เพื่อลดเวลาในการคำนวณ และหน่วยความจำเท่านั้น แต่ยังสามารถนำไปใช้แก้ปัญหาอื่นๆ ได้อีก ซึ่งก็ขึ้นกับว่าจะใช้เทคนิคใดในการเลือกโปรโตไทป์ โดยแต่ละเทคนิคก็จะช่วยแก้ปัญหาที่แตกต่างกันไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

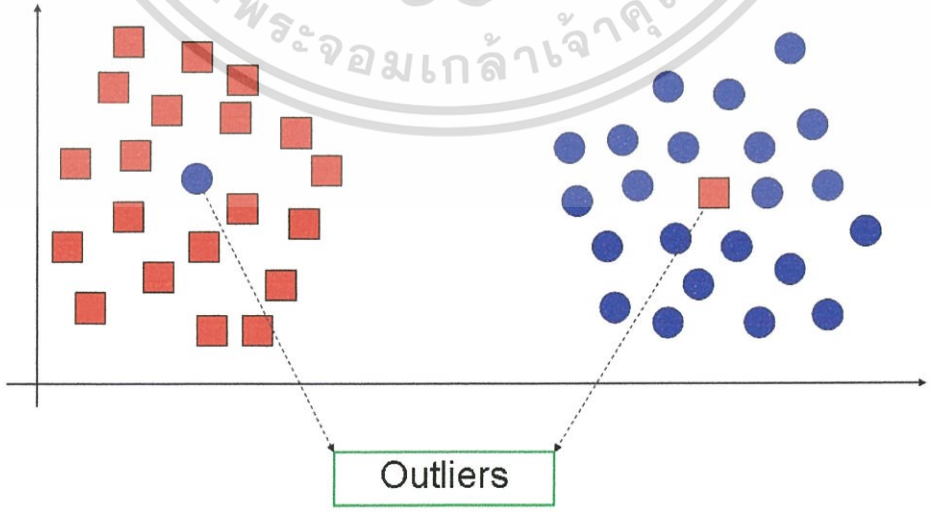
เทคนิคการเลือกโปรโตไทป์สามารถแบ่งตามจุดประสงค์ ได้ 2 ประเภท ได้ดังนี้ [6]

2.5.1 การเลือกโปรโตไทป์เพื่อลดจำนวนโปรโตไทป์ (Condensing Technique)

การเลือกโปรโตไทป์เพื่อลดจำนวนโปรโตไทป์เป็นเทคนิคของการเลือกโปรโตไทป์ที่มีเป้าหมายเพื่อลดจำนวนข้อมูลที่ใช้ในการอ้างอิงลงให้มีจำนวนน้อยที่สุดเท่าที่ทำได้ ในขณะที่ยังสามารถคงรักษาประสิทธิภาพในการจำแนกข้อมูลให้ใกล้เคียงกับการใช้ชุดข้อมูลฝึกสอนดั้งเดิมเป็นข้อมูลอ้างอิง เทคนิคนี้ใช้เพื่อลดเวลาในการคำนวณ และหน่วยความจำในการจำแนกข้อมูลโดยใช้ NN Rule ในกรณีที่มีโปรโตไทป์จำนวนมาก ตัวอย่างเทคนิคการเลือกโปรโตไทป์ในลักษณะนี้ เช่น วิธี MCS [4] และ GAs [5]

2.5.2 การเลือกโปรโตไทป์เพื่อกำจัดข้อมูลที่ไม่เข้าพวก (Editing Technique)

การเลือกโปรโตไทป์เพื่อกำจัดข้อมูลที่ไม่เข้าพวกเป็นเทคนิคของการเลือกโปรโตไทป์ที่มีเป้าหมายเพื่อลดข้อมูลที่ผิดพลาดที่แทรกอยู่ในชุดข้อมูลฝึกสอน หรือเพื่อลดข้อมูลที่มีลักษณะไม่เข้าพวก (Outlier) ตัวอย่างที่แสดงดังรูปที่ 2.5 คือ ข้อมูลที่วางตัวอยู่ระหว่างข้อมูลชนิดอื่น ไม่อยู่ในกลุ่มข้อมูลที่เป็นชนิดเดียวกัน ซึ่งเมื่อลดข้อมูลเหล่านี้ไปจนหมดแล้วจะได้กลุ่มของโปรโตไทป์ที่มีลักษณะเป็นกลุ่มก้อน โดยแต่ละกลุ่มจะเป็นโปรโตไทป์ที่มีชนิดเดียวกันรวมเข้าด้วยกัน และเรียกกลุ่มโปรโตไทป์ที่มีลักษณะแบบนี้ว่า "Well-Clustered Groups of Homogenous Prototype" เทคนิคนี้ช่วยลดข้อมูลที่ผิดพลาด (Noise Reduction) อีกทั้งยังสามารถช่วยลดปัญหาความจำเพาะของข้อมูล (Over Fitting) ได้ด้วย ซึ่งจะทำให้ประสิทธิภาพการจำแนกมีความถูกต้องมากขึ้น ตัวอย่างเทคนิคการเลือกโปรโตไทป์ในลักษณะนี้ เช่น Proximity Graphs [7]



รูปที่ 2.5 ตัวอย่างของข้อมูลที่ไม่เข้าพวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

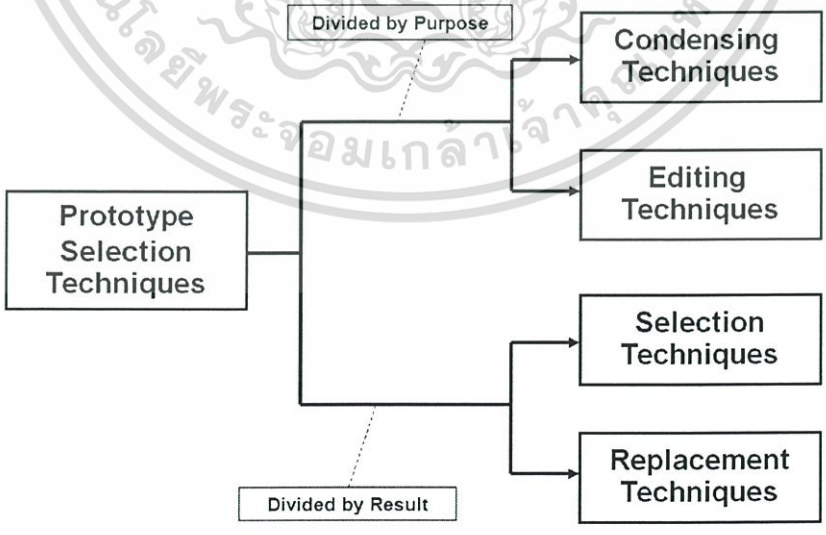
นอกจากการแบ่งเทคนิคการเลือกโปรโตไทป์ตามจุดประสงค์ของเทคนิคตามที่ได้กล่าวไปแล้ว เทคนิคการเลือกโปรโตไทป์ยังสามารถแบ่งตามลักษณะของผลลัพธ์ได้อีกด้วย โดยสามารถแบ่งได้เป็น 2 ประเภท ดังนี้ [6]

2.5.3 การเลือกโปรโตไทป์ที่ให้ผลลัพธ์ที่สร้างขึ้นใหม่ (Replacement Technique)

การเลือกโปรโตไทป์ที่ให้ผลลัพธ์ที่สร้างขึ้นใหม่เป็นเทคนิคการเลือกโปรโตไทป์ที่โปรโตไทป์ผลลัพธ์จะเป็นเซตย่อยของชุดฝึกสอนหรือไม่ก็ได้ หมายความว่าเทคนิคการเลือกโปรโตไทป์ในลักษณะนี้สามารถสร้างโปรโตไทป์ขึ้นมาใหม่ได้ ซึ่งถือเป็นข้อดีของวิธีนี้ เพราะโปรโตไทป์ผลลัพธ์ไม่จำเป็นต้องยึดติดกับเซตของชุดข้อมูลฝึกสอน จึงมีโอกาสที่จะได้โปรโตไทป์ผลลัพธ์ที่ดีกว่าวิธีการเลือกโปรโตไทป์ที่ให้ผลลัพธ์เป็นเซตย่อยชุดข้อมูลฝึกสอน ตัวอย่างเทคนิคการเลือกโปรโตไทป์ในลักษณะนี้ เช่น MCA [8] และ LVQ [9]

2.5.4 การเลือกโปรโตไทป์ที่ให้ผลลัพธ์เป็นเซตย่อย (Selection Technique)

การเลือกโปรโตไทป์ที่ให้ผลลัพธ์เป็นเซตย่อยเป็นเทคนิคการเลือกโปรโตไทป์ที่โปรโตไทป์ผลลัพธ์จะเป็นเซตย่อยของชุดข้อมูลฝึกสอนเสมอ หมายความว่าเทคนิคการเลือกโปรโตไทป์ในลักษณะนี้ก็คือการเลือกโปรโตไทป์จากชุดข้อมูลฝึกสอนเท่านั้น ดังนั้นเซตของโปรโตไทป์ผลลัพธ์จะเป็นเซตย่อยของชุดข้อมูลฝึกสอนเสมอ ตัวอย่างเทคนิคการเลือกโปรโตไทป์ในลักษณะนี้ เช่น CNN [10] และ MCS [4]



รูปที่ 2.6 แผนผังเทคนิคการเลือกโปรโตไทป์ แบบต่างๆ

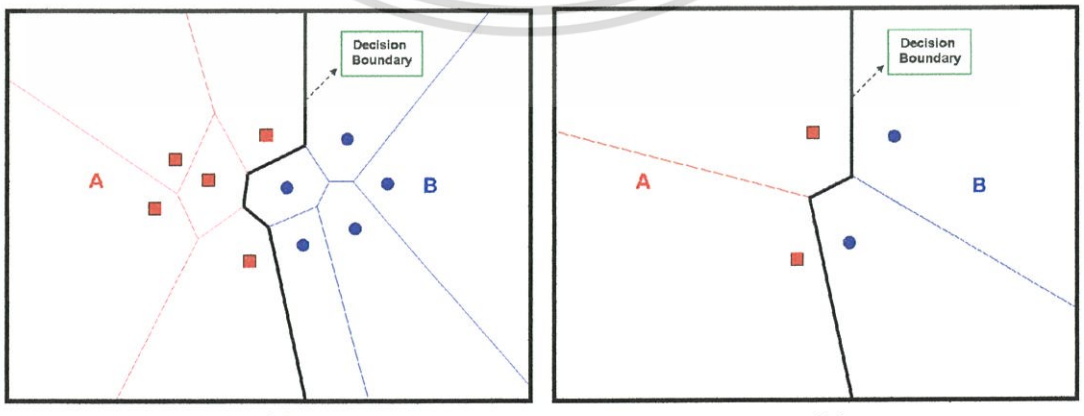
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเลือกโปรโตไทป์ในวิทยานิพนธ์ฉบับนี้จะเป็นแบบ “Condensing-Selection Prototype Selection” กล่าวคือเป็นการเลือกโปรโตไทป์เพื่อลดจำนวนข้อมูลอ้างอิงที่ใช้ในการจำแนกข้อมูล โดยใช้ NN Rule ลงโดยที่โปรโตไทป์ผลลัพธ์ที่ได้จะเป็นเซตย่อยของชุดข้อมูลฝึกสอน ในขณะที่ยังสามารถคงรักษาประสิทธิภาพในการจำแนกข้อมูลให้ใกล้เคียงกับการใช้ชุดข้อมูลฝึกสอนดั้งเดิมเป็นข้อมูลอ้างอิง โดยจะเรียกปัญหาแบบว่า “Minimal Consistent Set Problem” [11] ซึ่งเป็นปัญหาแบบ “NP” [12] กล่าวคือเป็นปัญหาที่ไม่อาจแก้ได้ด้วยทรัพยากร และเวลาที่จำกัด

จากหัวข้อที่ 2.5.1 จะเห็นว่าสิ่งสำคัญสิ่งหนึ่งในการเลือกโปรโตไทป์ในลักษณะนี้ คือ การจำแนกข้อมูลที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิงจะต้องมีประสิทธิภาพการจำแนกใกล้เคียงกับการจำแนกข้อมูลโดยใช้ชุดข้อมูลฝึกสอนเป็นข้อมูลอ้างอิง ซึ่งประสิทธิภาพการจำแนกข้อมูลนั้นสามารถแสดงได้ในรูปของขอบเขตการตัดสินใจ นั่นหมายความว่าการเลือกโปรโตไทป์ (เพื่อลดจำนวนข้อมูลอ้างอิง) ที่ดีนั้น โปรโตไทป์ผลลัพธ์ที่ได้จะต้องมีขอบเขตการตัดสินใจใกล้เคียงกับขอบเขตการตัดสินใจเดิม โดยปัญหาแบบนี้เรียกว่า “Minimal Consistent Set Problem”

2.6 ขอบเขตการตัดสินใจ (Decision Boundary)

ขอบเขตในการตัดสินใจของการจำแนกข้อมูลโดยใช้ NN Rule สามารถแสดงออกมาในรูปแบบที่เรียกว่า “Voronoi Diagram” [13] ตัวอย่างแสดงดังรูปที่ 2.7 โดยกำหนดให้ชุดข้อมูลฝึกสอนประกอบไปด้วยออบเจ็คชนิด A จำนวน 5 ออบเจ็คและออบเจ็คชนิด B จำนวน 5 ออบเจ็ค แสดงอยู่บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ จะเห็นว่าเส้นตรงสีดำกลางเป็นขอบเขตการตัดสินใจในแบ่งระหว่างชนิด A กับ ชนิด B นั่นหมายความว่าถ้ามีออบเจ็คไม่ทราบชนิดที่แสดงอยู่บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ วางตัวอยู่ทางซ้ายของขอบเขตการตัดสินใจ ตัวจำแนกก็จะจำแนกออบเจ็คไม่ทราบชนิดให้เป็นชนิด A ในทางกลับกันถ้าออบเจ็คไม่ทราบชนิดวางตัวอยู่ทางขวาของขอบเขตการตัดสินใจ ตัวจำแนกก็จะจำแนกออบเจ็คไม่ทราบชนิดให้เป็นชนิด B



(ก) (ข)

รูปที่ 2.7 ขอบเขตการตัดสินใจในรูปของ Voronoi Diagram

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

จากรูปที่ 2.7 แสดงขอบเขตการตัดสินใจในรูปของ Voronoi Diagram โดยรูป (ก) แสดงขอบเขตการตัดสินใจของชุดข้อมูลฝึกสอน ส่วนรูป (ข) แสดงขอบเขตการตัดสินใจของโปรโตไทป์ผลลัพธ์ที่ลดจำนวนข้อมูลอ้างอิงลงแล้ว ซึ่งจะเห็นว่าถ้าขอบเขตการตัดสินใจเปลี่ยนไป ความถูกต้องในการจำแนกก็จะเปลี่ยนไปด้วย ดังนั้นในการเลือกโปรโตไทป์ที่ดีนั้น โปรโตไทป์ผลลัพธ์ควรจะต้องมีขอบเขตการตัดสินใกล้เคียงกับขอบเขตการตัดสินใจเดิมให้มากที่สุด แต่การเปรียบเทียบเพื่อตรวจสอบขอบเขตการตัดสินใจโดยตรงนั้นทำได้ยาก จึงมีการนิยาม "คุณสมบัติความสอดคล้อง" ขึ้น

2.7 คุณสมบัติความสอดคล้อง (Consistency Property)

คุณสมบัติความสอดคล้องถูกนำเสนอไว้ในบทความที่มีชื่อว่า "The Condensed Nearest Neighbor Rule" [10] เป็นคุณสมบัติที่ใช้แทนการตรวจสอบขอบเขตการตัดสินใจของโปรโตไทป์ผลลัพธ์ที่ได้จากการเลือกโปรโตไทป์ว่ามีขอบเขตการตัดสินใจใกล้เคียงเดิมหรือไม่

ขั้นตอนการตรวจสอบสามารถทำได้ง่ายๆ โดยการใช้โปรโตไทป์ผลลัพธ์ข้อมูลอ้างอิงในการจำแนกข้อมูลโดยใช้ NN Rule โดยถ้าโปรโตไทป์ผลลัพธ์สามารถจำแนกข้อมูลชุดข้อมูลฝึกสอนได้ถูกต้องทุกตัว ก็จะหมายความว่าขอบเขตการตัดสินใจของโปรโตไทป์ผลลัพธ์มีความใกล้เคียงกับขอบเขตการตัดสินใจของชุดข้อมูลฝึกสอน โดยจะเรียกว่า "โปรโตไทป์ผลลัพธ์มีความสอดคล้องกับชุดข้อมูลฝึกสอน"

นิยาม ความสอดคล้อง (Consistency Property)
"โปรโตไทป์ผลลัพธ์จากการเลือกโปรโตไทป์จะมีความสอดคล้องกับชุดข้อมูลฝึกสอนก็ต่อเมื่อการจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิง สามารถจำแนกข้อมูลในชุดข้อมูลฝึกสอนได้ถูกต้องทุกตัว" [10]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
105133

บทที่ 3
งานวิจัยที่เกี่ยวข้อง

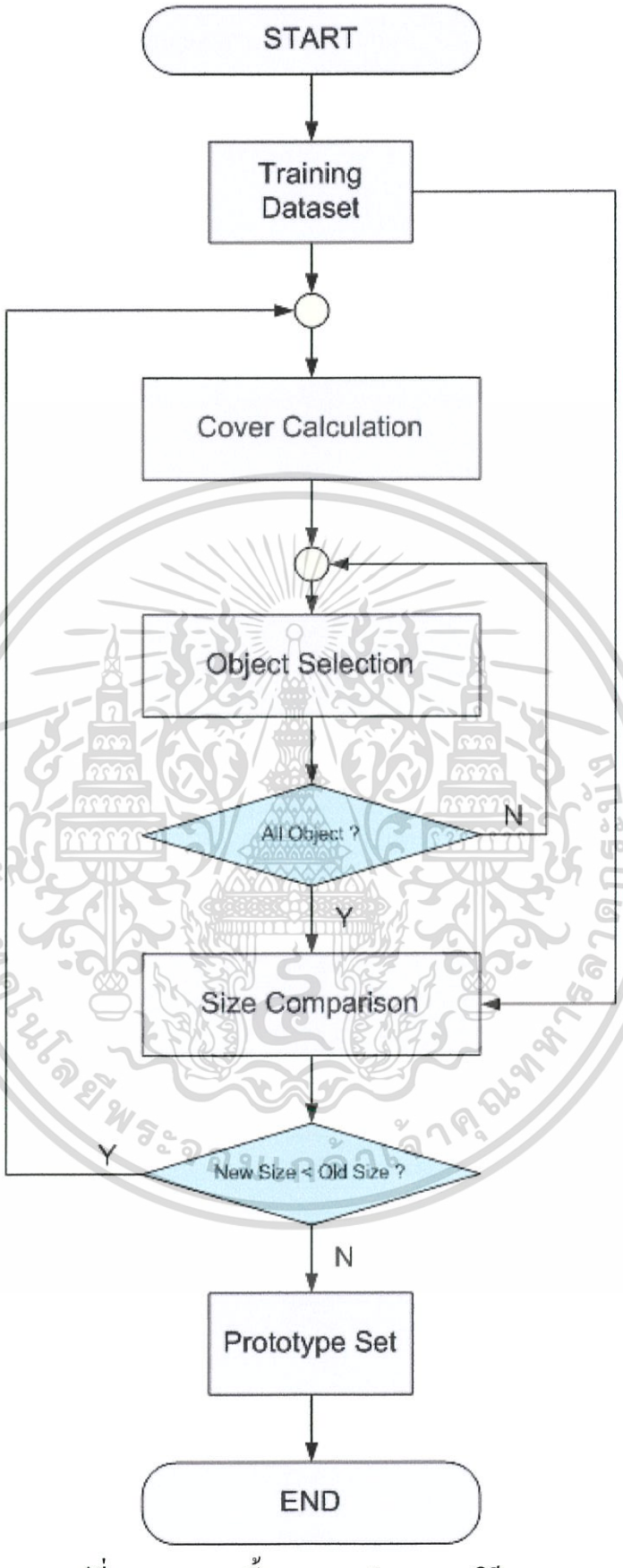
3.1 Minimal Consistent Set Identification (MCS)

วิธี MCS ถูกนำเสนอโดย Belur V. Dasarathy ในบทความที่มีชื่อว่า "Minimal Consistent Set (MCS) Identification for Optimal Nearest Neighbor Decision Systems Design" [4] ซึ่งเป็นการเลือกโปรโตไทป์แบบ "Condensing-Selection" ที่อาศัยหลักการ "ความครอบคลุม" ในการทำงาน กล่าวคือเป็นการเลือกโปรโตไทป์โดยอาศัยความรู้ที่ได้จากการคำนวณความครอบคลุมในการกรองออบเจ็คที่ซ้ำซ้อนออก ทำให้ได้เฉพาะข้อมูลที่จำเป็นเท่านั้น อีกทั้งโปรโตไทป์ผลลัพธ์ที่ได้ยังรับประกันว่ามีความสอดคล้องกับชุดข้อมูลฝึกสอนอีกด้วย

ขั้นตอนการทำงานของวิธี MCS สามารถอธิบายได้ง่ายๆ คือ การให้แต่ละออบเจ็คทำการลงคะแนน (Vote) เพื่อเฟ้นหาตัวแทนของออบเจ็คนั้นๆ โดยตัวแทนก็คือออบเจ็คที่สามารถใช้เป็นข้อมูลอ้างอิงในการจำแนกข้อมูลโดยใช้ NN Rule แล้วสามารถจำแนกออบเจ็คที่ถูกแทนได้ถูกต้อง และเมื่อทำการลงคะแนนครบทุกออบเจ็คแล้ว ก็จะทำการนับผลคะแนนเพื่อพิจารณาออบเจ็คที่มีผลคะแนนสูงสุด ซึ่งหมายความว่าออบเจ็คนั้นเป็นตัวแทนที่สามารถใช้แทนออบเจ็คอื่นๆ ได้มากที่สุด จึงควรเลือกออบเจ็คนี้เป็นโปรโตไทป์ และเมื่อทำการเลือกออบเจ็คที่มีผลคะแนนมากที่สุดเป็นโปรโตไทป์แล้วก็หมายความว่าสามารถลดออบเจ็คที่ถูกแทนแล้วลงได้

วิธี MCS เป็นการค้นหาแบบเฉพาะที่ (Local Search) ทำงานแบบ "Bottom Up" กล่าวคือจะทำงานเป็นรอบๆ โดยในรอบแรกจะเริ่มต้นจากนำออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนมาใช้ในการคำนวณความครอบคลุม จากนั้นนำความรู้ที่ได้จากการคำนวณมาทำการเลือกโปรโตไทป์ ส่วนในรอบที่สองก็จะนำโปรโตไทป์ผลลัพธ์ที่ได้ในรอบแรกมาใช้ในการคำนวณความครอบคลุม และทำการเลือกโปรโตไทป์อีกที โดยจะวนทำงานไปเรื่อยๆ จนกว่าจะไม่สามารถเลือกโปรโตไทป์ที่มีจำนวนน้อยลงกว่าเดิมได้อีก ดังแสดงแผนภาพขั้นตอนการทำงานของวิธี MCS ในรูปที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



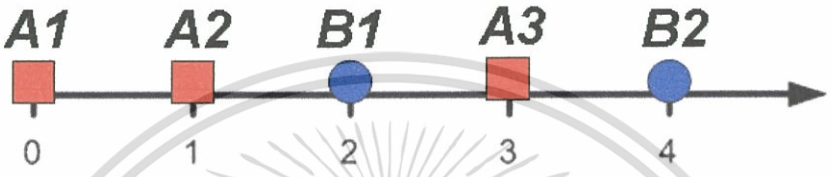
รูปที่ 3.1 แผนภาพขั้นตอนการทำงานของวิธี MCS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 นิยามที่ควรทราบเกี่ยวกับวิธี MCS

3.1.1.1 ออบเจ็คต่างชนิดที่อยู่ใกล้ที่สุด (Nearest Unlike Neighbor)

ออบเจ็คต่างชนิดที่อยู่ใกล้ที่สุด (เรียกสั้นๆ ว่า NUN) เป็นออบเจ็คที่อยู่ใกล้กับออบเจ็คที่กำลังพิจารณามากที่สุด และมีชนิดต่างกับออบเจ็คที่กำลังพิจารณา ตัวอย่างเช่น กำหนดให้ชุดข้อมูลตัวอย่างที่ 1 ประกอบไปด้วยออบเจ็คชนิด A จำนวน 3 ออบเจ็ค (ออบเจ็ค A1, ออบเจ็ค A2 และออบเจ็ค A3) และออบเจ็คชนิด B จำนวน 2 ออบเจ็ค (ออบเจ็ค B1 และออบเจ็ค B2) แสดงอยู่บนปริภูมิคุณลักษณะเฉพาะ 1 มิติ ดังแสดงดังรูปที่ 3.2



รูปที่ 3.2 ชุดข้อมูลตัวอย่างที่ 1 บนปริภูมิคุณลักษณะเฉพาะ 1 มิติ

จากรูปที่ 3.2 แสดงชุดข้อมูลตัวอย่างที่ 1 บนปริภูมิคุณลักษณะเฉพาะ 1 มิติ เมื่อพิจารณาที่ออบเจ็ค A1 จะเห็นว่าออบเจ็ค A1 อยู่ห่างจากออบเจ็ค B1 เท่ากับ "2" หน่วย และอยู่ห่างจากออบเจ็ค B2 เท่ากับ "4" หน่วย ซึ่งจะเห็นว่าออบเจ็ค A1 นั้นอยู่ใกล้กับออบเจ็ค B1 มากกว่าออบเจ็ค B2 และออบเจ็ค B1 มีชนิดต่างกับออบเจ็ค A1 ดังนั้นจะเรียกว่า "ออบเจ็ค B1 เป็นออบเจ็คต่างชนิดที่อยู่ใกล้ออบเจ็ค A1 มากที่สุด" หรือ "ออบเจ็ค B1 เป็น NUN ของออบเจ็ค A1"

<u>นิยาม</u> ออบเจ็คต่างชนิดที่อยู่ใกล้ที่สุด (NUN)

"ออบเจ็ค A เป็นออบเจ็คต่างชนิดที่อยู่ใกล้ที่สุดของออบเจ็ค B ก็ต่อเมื่อ ออบเจ็ค A กับออบเจ็ค B มีชนิดที่แตกต่างกัน และระยะห่างระหว่างออบเจ็ค A กับออบเจ็ค B มีค่าน้อยที่สุดเมื่อเทียบกับออบเจ็คอื่นๆ"

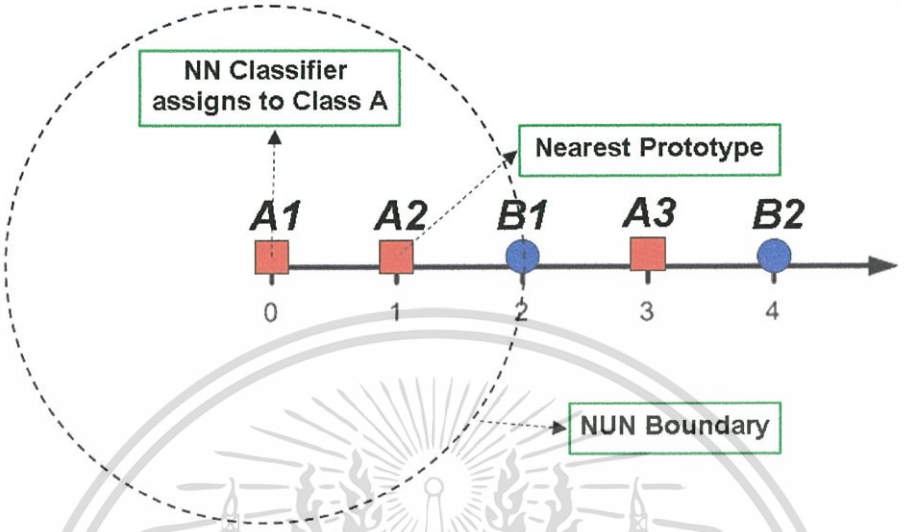
3.1.1.2 รัศมีขอบเขตต่างชนิด (NUN Boundary)

รัศมีขอบเขตต่างชนิดเป็นรัศมีจากออบเจ็คที่กำลังพิจารณาไปยัง NUN ของออบเจ็คนั้น โดยภายในขอบเขตจะมีแต่ออบเจ็คที่มีชนิดเดียวกับออบเจ็คที่กำลังพิจารณาเท่านั้น ซึ่งออบเจ็คเหล่านี้จะสามารถให้เป็นข้อมูลอ้างอิงในการจำแนกออบเจ็คที่กำลังพิจารณาโดยใช้ NN Rule ได้ถูกต้อง ตัวอย่างจะแสดงในหัวข้อถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1.3 ความครอบคลุม (Cover)

ความครอบคุลม คือ การที่ออบเจ็คที่กำลังพิจารณาสามารถใช้เป็นข้อมูลอ้างอิงในการจำแนกออบเจ็คใดๆ โดยใช้ NN Rule ได้ถูกต้อง ดังแสดงตัวอย่างในรูปที่ 3.3



รูปที่ 3.3 การจำแนกชนิดของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 1 โดยใช้ NN Rule

จากรูปที่ 3.3 แสดงการจำแนกชนิดของออบเจ็ค A1 โดยใช้ NN Rule ซึ่งจะเห็นว่าถ้าต้องการจำแนกออบเจ็ค A1 ให้ถูกต้อง ก็จะต้องเลือกออบเจ็ค A2 ที่เป็นออบเจ็คชนิดเดียวกันกับออบเจ็ค A1 เป็นโปรโตไทป์ (เพราะในการจำแนกข้อมูลโดยใช้ NN Rule จะกำหนดให้ออบเจ็คที่ต้องการทราบชนิดเป็นชนิดเดียวกันกับโปรโตไทป์ที่อยู่ใกล้กับออบเจ็คนั้นๆ มากที่สุด) ดังนั้นจะเรียกว่า "ออบเจ็ค A2 ครอบคลุมออบเจ็ค A1" (หรือก็คือสามารถใช้ออบเจ็ค A2 เป็นตัวของแทนออบเจ็ค A1 ได้) และจะเห็นว่าออบเจ็ค A3 ไม่ครอบคลุมออบเจ็ค A1 เพราะเมื่อทำการจำแนกชนิดของออบเจ็ค A1 โดยใช้ NN Rule จะจำแนกได้ชนิด B เพราะออบเจ็คที่อยู่ใกล้กับออบเจ็ค A1 มากที่สุด คือ ออบเจ็ค B1 ซึ่งเป็นการจำแนกข้อมูลที่ผิด

อีกทางเลือกหนึ่งในการเลือกโปรโตไทป์ก็คือ ตัวออบเจ็ค A1 เอง ซึ่งแน่นอนว่าถ้าเลือกออบเจ็ค A1 เป็นโปรโตไทป์ก็จะสามารถใช้เป็นข้อมูลในการอ้างอิงในการจำแนกออบเจ็ค A1 ได้ถูกต้องแน่นอน นั่นหมายความว่าไม่ว่าจะเลือกออบเจ็คใดที่อยู่ภายในรัศมีขอบเขตต่างชนิด ของออบเจ็ค A1 ก็จะสามารถใช้เป็นโปรโตไทป์ในการจำแนกออบเจ็ค A1 โดยใช้ NN Rule ได้ถูกต้อง

<u>นิยาม</u> ความครอบคลุม

"ออบเจ็ค A จะครอบคลุมออบออบเจ็ค B ก็ต่อเมื่อ การจำแนกข้อมูลโดยใช้ NN Rule ที่มีออบเจ็ค A เป็นข้อมูลอ้างอิง สามารถจำแนกออบเจ็ค B ได้ถูกต้อง"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1.4 ค่าความสามารถในการครอบคลุม (Cover Value)

ค่าความสามารถในการครอบคลุม คือ จำนวนออบเจ็คของออบเจ็คที่กำลังพิจารณาสามารถครอบคลุมได้ (หรือก็คือจำนวนออบเจ็คที่สามารถใช้ออบเจ็คที่กำลังพิจารณาเป็นตัวแทนได้) ซึ่งเมื่อทำการเลือกโปรโตไทป์ตัวแทนแล้ว ออบเจ็คที่มีตัวแทนแล้วเหล่านั้นก็ไม่มีความจำเป็นอีกต่อไป เพราะออบเจ็คเหล่านั้นสามารถถูกแทนได้โดยตัวแทนแล้ว ทำให้สมารถลดออบเจ็คที่มีตัวแทนแล้วเหล่านั้นลงได้

3.1.1.5 เซตของออบเจ็คที่น่าจะเป็นโปรโตไทป์ (Candidate Set)

เซตของออบเจ็คที่น่าจะเป็นโปรโตไทป์ (Candidate Set) คือ เซตของออบเจ็คที่มีแนวโน้มที่ดีที่น่าจะเป็นโปรโตไทป์ ซึ่งจะใช้เป็นอินพุตของวิธี MCS ในการทำงานแต่ละรอบ กล่าวคือจะทำการเลือกออบเจ็คจาก Candidate Set เป็นโปรโตไทป์ โดยในการทำงานรอบแรกที่ยังไม่ทราบว่าออบเจ็คใดบ้างจะถูกเลือกเป็นโปรโตไทป์ ดังนั้นจึงกำหนดให้ทุกออบเจ็คมีโอกาสถูกเลือกเป็นโปรโตไทป์ได้ นั่นคือ Candidate Set ในการทำงานรอบแรกก็คือออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนนั่นเอง

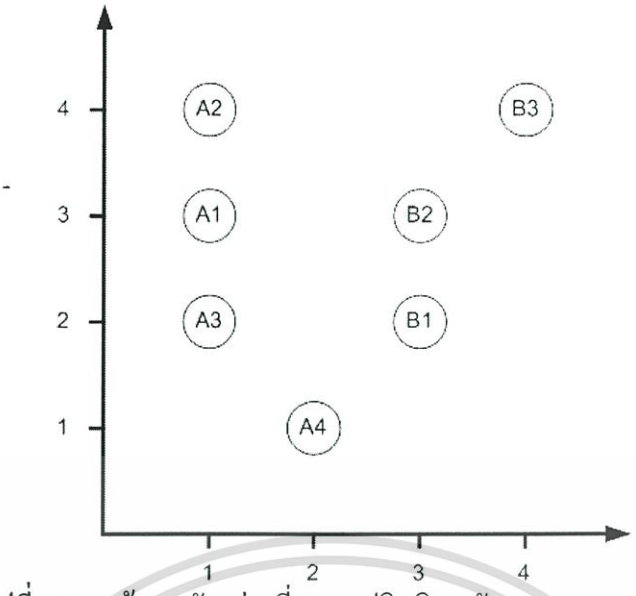
3.1.2 รายละเอียดขั้นตอนการทำงานของวิธี MCS

3.1.2.1 กระบวนการคำนวณระยะทาง (Distance Calculating Process)

กระบวนการสร้างตารางระยะทางเป็นกระบวนที่ทำการคำนวณระยะห่างระหว่าง 2 ออบเจ็คที่เป็นไปได้ทั้งหมดในชุดข้อมูลฝึกสอน โดยใช้ฟังก์ชั่นคำนวณระยะทาง จากนั้นนำค่าระยะห่างที่ได้มาเก็บไว้ในตารางระยะทาง เพื่อนำไปใช้ในการคำนวณความครอบคุลมต่อไป โดยกระบวนนี้จะทำงานเพียงครั้งเดียว เพราะถ้าต้องการทราบระยะห่างระหว่างออบเจ็คใดๆ ก็สามารถมาพิจารณาจากตารางระยะทางได้เลย

ตัวอย่างเช่น กำหนดให้ชุดข้อมูลตัวอย่างที่ 2 ประกอบไปด้วยออบเจ็คจำนวน 7 ออบเจ็ค ซึ่งประกอบไปด้วยออบเจ็คชนิด A จำนวน 4 ตัว และออบเจ็คชนิด B จำนวน 3 ตัว ดังนี้ A1 = ((1,3),A), A2 = ((1,4),A), A3 = ((1,2),A), A4 = ((2,1),A), B1 = ((3,2),B), B2 = ((3,3),B), B3 = ((4,4),B) ดังแสดงอยู่บนปริภูมิคุณลักษณะเฉพาะ 2 มิติในรูปที่ 3.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 ชุดข้อมูลตัวอย่างที่ 2 บนปริภูมิคุณลักษณะเฉพาะ 2 มิติ

เมื่อนำออบเจ็คทั้งหมดในชุดข้อมูลตัวอย่างที่ 2 มาคำนวณหาระยะทางระหว่าง 2 ออบเจ็คที่เป็นไปได้ทั้งหมดโดยใช้การวัดระบบแบบยูคลิด ดังแสดงในตารางที่ 3.1

ตารางที่ 3.1 ข้อมูลระยะทางของชุดข้อมูลตัวอย่างที่ 2

	A1	A2	A3	A4	B1	B2	B3
A1	0.00	1.00	1.00	2.23	2.23	2.00	3.16
A2	1.00	0.00	2.00	3.16	2.82	2.23	3.00
A3	1.00	2.00	0.00	1.41	2.00	2.23	3.60
A4	2.23	3.16	1.41	0.00	1.41	2.23	3.60
B1	2.23	2.82	2.00	1.41	0.00	1.00	2.23
B2	2.00	2.23	2.23	2.23	1.00	0.00	1.41
B3	3.16	3.00	3.60	3.60	2.23	1.41	0.00

3.1.2.2 กระบวนการคำนวณความครอบคลุม (Cover Calculating Process)

กระบวนการสร้างตารางความครอบคลุมเป็นกระบวนการที่จะคำนวณความครอบคลุม เพื่อใช้ในการเลือกโปรโตไทป์ในขั้นตอนถัดไป โดยพิจารณาหาว่าแต่ละออบเจ็คนั้น มีออบเจ็คใดบ้างที่สามารถครอบคลุมออบเจ็คนั้นๆ ได้ (หรือก็คือพิจารณาว่าแต่ละออบเจ็คมีออบเจ็คให้บ้างที่สามารถเป็นตัวแทนออบเจ็คนั้นๆ ได้) โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **การเรียงลำดับระยะทาง (Distance Sorting)**

การเรียงลำดับระยะทางเป็นการนำข้อมูลในตารางระยะทางของแต่ละออบเจ็คมาเรียงลำดับจากน้อยไปหามาก ดังแสดงตัวอย่างในรูปที่ 3.5 และเมื่อทำการเรียงลำดับข้อมูลในตารางระยะทางครบทุกออบเจ็คแล้ว ดังแสดงในตารางที่ 3.2

	A1	A2	A3	A4	B1	B2	B3
A1	0.00	1.00	1.00	2.23	2.23	2.00	3.16

Sorting by Distance

	A1	A2	A3	B2	B1	A4	B3
A1	0.00	1.00	1.00	2.00	2.23	2.23	3.16

รูปที่ 3.5 การเรียงลำดับข้อมูลของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2

ตารางที่ 3.2 ข้อมูลระยะทางของชุดข้อมูลตัวอย่างที่ 2 ที่เรียงลำดับแล้ว

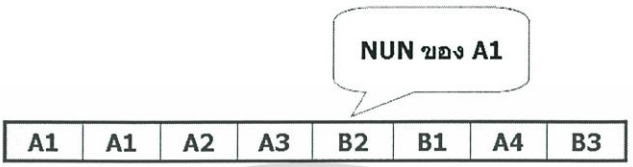
A1	A1	A2	A3	B2	B1	A4	B3
A2	A2	A1	A3	B2	B1	B3	A4
A3	A3	A1	A4	B1	A2	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

- **การเลือกออบเจ็คต่างชนิด (NUN Selection)**

การเลือกออบเจ็คต่างชนิดเป็นการนำข้อมูลในตารางระยะทางที่เรียงลำดับแล้วมาพิจารณาเลือก NUN เพื่อใช้ในการพิจารณารัศมีขอบเขตต่างชนิดของแต่ละออบเจ็ค โดยในการเลือก NUN นั้นจะต้องพิจารณาจาก Candidate Set เท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากชุดข้อมูลตัวอย่างที่ 2 ซึ่งประกอบไปด้วยออบเจ็ค A1, A2, A3, A4, B1, B2 และ B3 ดังนั้น Candidate Set ของการทำงานรอบแรกก็คือ {A1, A2, A3, A4, B1, B2, B3} จากนั้นทำการพิจารณาข้อมูลระยะทางที่เรียงลำดับแล้วของแต่ละออบเจ็คจากใกล้ไปไกล (หรือก็คือพิจารณาจากซ้ายไปขวา) และเมื่อเจอออบเจ็คที่มีชนิดต่างกับออบเจ็คที่กำลังพิจารณา ก็จะทำการเลือกออบเจ็คนั้นเป็น NUN ของออบเจ็คที่กำลังพิจารณา ดังแสดงตัวอย่างในรูปที่ 3.6 และเมื่อทำการพิจารณาเลือก NUN ในรอบแรกครบทุกออบเจ็คแล้ว สามารถแสดงได้ดังในตารางที่ 3.3



รูปที่ 3.6 การเลือก NUN ในรอบแรกของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2

จากรูปที่ 3.6 แสดงการเลือก NUN ของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 ที่นำข้อมูลของออบเจ็ค A1 จากตางที่ 3.2 มาพิจารณาจากใกล้ไปไกล (ซ้ายไปขวา) ซึ่งจะเห็นว่า B2 อยู่ใน Candidate Set และอยู่ใกล้กับออบเจ็ค A1 มากที่สุด และเป็นมีชนิดต่างกันกับออบเจ็ค A1 ดังนั้นจึงทำการเลือกออบเจ็ค B2 เป็น NUN ของออบเจ็ค A1

ตารางที่ 3.3 การเลือก NUN ในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2

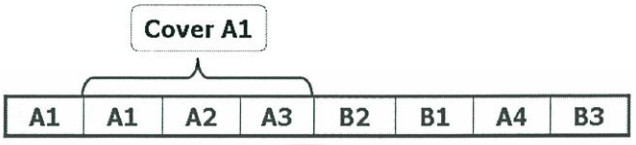
Candidate Set = {A1, A2, A3, A4, B1, B2, B3}

A1	A1	A2	A3	B2	B1	A4	B3
A2	A2	A1	A3	B2	B1	B3	A4
A3	A3	A1	A4	B1	A2	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **การพิจารณาความครอบคลุม (Cover Determination)**

การพิจารณาความครอบคลุมเป็นการพิจารณาออบเจ็คที่สามารถครอบคลุมแต่ละออบเจ็คได้ (หรือก็คือพิจารณาออบเจ็คที่สามารถเป็นตัวแทนแต่ละออบเจ็คได้) ซึ่งสามารถทำได้โดยการพิจารณารัศมีขอบเขตต่างชนิด โดยถ้าออบเจ็คใดอยู่ในรัศมีขอบเขตต่างชนิดก็หมายความว่าออบเจ็คเหล่านั้นสามารถครอบคลุมออบเจ็คที่กำลังพิจารณาได้ (ตามที่อธิบายไปในหัวข้อที่ 3.1.1.3) ดังแสดงตัวอย่างในตารางที่ 3.7



รูปที่ 3.7 ออบเจ็คที่สามารถครอบคุลมออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2 ได้

จากรูปที่ 3.7 แสดงออบเจ็คที่สามารถครอบคุลมออบเจ็ค A1 ได้ ซึ่งจะเห็นว่าออบเจ็ค B2 เป็น NUN ของออบเจ็ค A1 ซึ่งจะเห็นว่าออบเจ็ค A1, A2 และ A3 อยู่ในรัศมีขอบเขตต่างชนิดของออบเจ็ค A1 นั่นหมายความว่าออบเจ็ค A1, A2 และ A3 ครอบคลุมออบเจ็ค A1

เมื่อทำการพิจารณาความครอบคลุมแล้วก็ทำการสร้างตารางความครอบคลุม โดยถ้าออบเจ็คใดสามารถครอบคลุมออบเจ็คที่กำลังพิจารณาได้ ก็จะใส่ค่า “1” ในช่องออบเจ็คนั้น ส่วนถ้าไม่สามารถครอบคลุมได้ก็จะใส่ค่า “0” ในช่องออบเจ็คนั้น ดังแสดงตัวอย่างในรูปที่ 3.8 และเมื่อทำการพิจารณาความครอบคลุมครบทุกออบเจ็คแล้วก็สามารถแสดงได้ดังในตารางที่ 3.4

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0

รูปที่ 3.8 ส่วนหนึ่งของตารางความรอบคลุมในรอบแรกของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 ตารางความครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2

Candidate Set = {A1, A2, A3, A4, B1, B2, B3}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	1	0	0	0	0
A3	1	0	1	1	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	1	1	0
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1

- **การคำนวณค่าความสามารถในการครอบคลุม (Cover Value Calculation)**

การคำนวณค่าความสามารถในการครอบคลุมเป็นพิจารณาว่าแต่ละออบเจ็คสามารถครอบคลุมออบเจ็คในชุดข้อมูลฝึกสอนได้มากน้อยเพียงใด (หรือก็คือพิจาณาว่าแต่ละออบเจ็คสามารถนำไปเป็นตัวแทนออบเจ็คในชุดข้อมูลฝึกสอนได้มากน้อยเพียงใด) ซึ่งสามารถทำได้โดยการพิจารณาข้อมูลความครอบคลุมของแต่ละออบเจ็ค และเรียกจำนวนนี้ว่า "ค่าความสามารถในการครอบคลุม (Cover Value : CV)" ดังแสดงในตารางที่ 3.5

ตารางที่ 3.5 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2

Candidate Set = {A1, A2, A3, A4, B1, B2, B3}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	1	0	0	0	0
A3	1	0	1	1	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	1	1	0
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1
CV	3	2	3	2	3	3	2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2.3 กระบวนการเลือกออบเจ็ค (Object Selecting Process)

กระบวนการเลือกออบเจ็คเป็นกระบวนการที่ทำการเลือกออบเจ็คเป็นโปรโตไทป์ ตัวอย่างเช่น เมื่อพิจารณาออบเจ็ค A1 ในตารางที่ 3.5 จะเห็นว่าออบเจ็คที่สามารถครอบคลุมออบเจ็ค A1 ได้มีออบเจ็ค A1, A2 และ A3 นั่นคือถ้าเลือกออบเจ็คใดออบเจ็คหนึ่งจาก 3 ออบเจ็คนี้เป็นโปรโตไทป์ก็จะสามารถจำแนกออบเจ็ค A1 โดยใช้ NN Rule ได้ถูกต้อง (หรือก็คือออบเจ็ค A1 สามารถถูกแทนได้โดยออบเจ็คใดออบเจ็คหนึ่งจาก 3 ออบเจ็คนั้น) และเมื่อทำการเลือกออบเจ็คใดออบเจ็คหนึ่งจาก 3 ออบเจ็คเป็นโปรโตไทป์แล้ว ออบเจ็ค A1 ก็ไม่มีความจำเป็นอีกต่อไป เพราะออบเจ็ค A1 สามารถถูกจำแนกโดยใช้ NN Rule ได้ถูกต้องโดยออบเจ็คใดออบเจ็คหนึ่งใน 3 ออบเจ็คนั้นแล้ว

แต่การจะเลือกออบเจ็ค A1, A2 หรือ A3 เป็นโปรโตไทป์นั้นไม่ได้ขึ้นกับออบเจ็ค A1 เพียงออบเจ็คเดียวแต่ต้องพิจารณาจากทุกออบเจ็ค ดังนั้นจึงต้องทำการพิจารณาจาก "ค่าความสามารถในการครอบคลุม" ซึ่งเป็นค่าที่บอกว่าแต่ละออบเจ็คสามารถครอบคลุมออบเจ็คในชุดข้อมูลฝึกสอนได้มากน้อยเพียงใด และสิ่งสำคัญอีกสิ่งหนึ่งในการเลือกออบเจ็คก็คือจะต้องทำการเลือกออบเจ็คจาก Candidate Set เท่านั้น

เมื่อพิจารณาตารางที่ 3.5 จะเห็นว่าออบเจ็ค A1, A3, B2 และ B3 อยู่ใน Candidate Set และมีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ "3" นั่นหมายความว่าถ้าเลือกออบเจ็คใดออบเจ็คหนึ่งจาก 4 ออบเจ็คนี้เป็นโปรโตไทป์ก็จะสามารถจำแนกออบเจ็คในชุดข้อมูลตัวอย่างที่ 2 โดยใช้ NN Rule ได้ถูกต้องจำนวน 3 ออบเจ็ค (หรือก็คือใช้ออบเจ็คเพียงตัวเดียวเป็นตัวแทนออบเจ็คในชุดข้อมูลตัวอย่างที่ 2 ได้ 3 ออบเจ็ค)

โดยตัวอัลกอริธึมดั้งเดิมของ MCS ไม่ได้กล่าวถึงวิธีการเลือกออบเจ็คในกรณีที่ออบเจ็คมีค่าความสามารถในการครอบคลุมสูงสุดเท่ากันไว้ ดังนั้นจะเลือกออบเจ็ค A1, A3, B2 หรือ B3 ก็ได้ โดยในที่นี้สมมุติว่าเลือกออบเจ็ค A1 เป็นโปรโตไทป์

เมื่อทำการเลือกออบเจ็ค A1 เป็นโปรโตไทป์แล้วจะเห็นว่าออบเจ็ค A1 ครอบคลุมออบเจ็ค A1, A2 และ A3 หมายความว่าสามารถใช้โปรโตไทป์ A1 เพียงตัวเดียวเป็นตัวแทนออบเจ็ค A1, A2 และ A3 ได้ กล่าวคือออบเจ็คเหล่านี้จึงไม่มีความจำเป็นอีก ดังนั้นค่าความสามารถในการครอบคลุมของออบเจ็ค A1, A2 และ A3 จะถูกลดทอน (ไม่นำมาคิดอีก) ออก จากนั้นจะทำการคำนวณค่าความสามารถในการครอบคลุมใหม่ ดังแสดงในตารางที่ 3.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 เป็นโปรโตไทป์แล้ว

Candidate Set = {A1, A2, A3, A4, B1, B2, B3}

Prototype Set = {A1}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	X	0	0	0	0
A3	X	0	X	X	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	1	1	0
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1
CV	0	0	0	1	3	3	2

เมื่อพิจารณาตารางที่ 3.6 จะเห็นว่าค่าความสามารถในการครอบคลุมเปลี่ยนไป กล่าวคือจะไม่นำค่าความสามารถในการครอบคลุมของออบเจ็ค A1, A2 และ A3 มาคิดอีก ซึ่งจะเห็นว่าค่าความสามารถในการครอบคลุมของออบเจ็ค A1, A2 และ A3 จะเท่ากับ "0" เพราะออบเจ็คเหล่านี้ถูกครอบคลุมโดยออบเจ็ค A1 ที่ถูกเลือกเป็นโปรโตไทป์ไปแล้ว ออบเจ็คเหล่านี้จึงไม่มีความจำเป็นอีกต่อไป อีกทั้งจะเห็นว่าค่าความสามารถในการครอบคลุมของออบเจ็ค A4 เปลี่ยนไป จากเดิมเท่ากับ "2" เพราะเดิมออบเจ็ค A4 ครอบคลุมออบเจ็ค A3 และออบเจ็ค A4 แต่ในเมื่อ A3 ถูกครอบคลุมไปแล้ว จึงไม่มีความจำเป็นที่จะต้องครอบคลุมออบเจ็ค A3 อีก ดังนั้นค่าความสามารถในการครอบคลุมจึงถูกลดทอนลงเหลือเท่ากับ "1"

จากนั้นทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจาก Candidate Set ลำดับต่อไปเป็นโปรโตไทป์ ซึ่งเมื่อพิจารณาตารางที่ 3.6 จะเห็นว่าออบเจ็ค B1 และ B2 อยู่ใน Candidate Set และมีค่าความสามารถในการครอบคลุมมากที่สุดเท่ากับ "3" ดังนั้นจะเลือกออบเจ็ค B1 หรือออบเจ็ค B2 ก็ได้ โดยในที่นี้สมมุติว่าเลือก B2 เป็นโปรโตไทป์

เมื่อทำการเลือกออบเจ็ค A1 เป็นโปรโตไทป์แล้วจะเห็นว่าออบเจ็ค B2 ครอบคลุมออบเจ็ค B1, B2 และ B3 ดังนั้นค่าความสามารถในการครอบคลุมของออบเจ็ค B1, B2 และ B3 จะถูกลดทอนออก จากนั้นจะทำการคำนวณค่าความสามารถในการครอบคลุมใหม่ ดังแสดงในตารางที่ 3.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.7 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 และ B2 เป็นโปรโตไทป์แล้ว

Candidate Set = {A1, A2, A3, A4, B1, B2, B3}

Prototype Set = {A1, B2}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	X	0	0	0	0
A3	X	0	X	X	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	X	X	0
B2	0	0	0	0	X	X	X
B3	0	0	0	0	X	X	X
CV	0	0	0	1	0	0	0

ขั้นตอนสุดท้ายก็ทำการเลือกออบเจ็ค A4 ที่เหลืออยู่เพียงตัวเดียวเป็นโปรโตไทป์ และเมื่อทำการเลือกออบเจ็ค A4 เป็นโปรโตไทป์แล้ว ค่าความสามารถในการครอบคลุมของออบเจ็ค A4 จะถูกลดทอนออก ดังแสดงในตารางที่ 3.8

ตารางที่ 3.8 ค่าความสามารถในการครอบคลุมในการทำงานรอบแรกของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1, B2 และ A4 เป็นโปรโตไทป์แล้ว

Candidate Set = {A1, A2, A3, A4, B1, B2, B3}

Prototype Set = {A1, B2, A4}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	X	0	0	0	0
A3	X	0	X	X	0	0	0
A4	0	0	0	X	0	0	0
B1	0	0	0	0	X	X	0
B2	0	0	0	0	X	X	X
B3	0	0	0	0	X	X	X
CV	0	0	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อพิจารณาตารางที่ 3.8 จะเห็นว่าค่าความสามารถในการครอบคลุมของทุกออบเจ็คเท่ากับ "0" เพราะทุกออบเจ็คในชุดข้อมูลตัวอย่างที่ 2 ถูกครอบคลุมโดยโปรโตไทป์ที่เลือกแล้ว (หรือก็คือทุกออบเจ็คในชุดข้อมูลตัวอย่างที่ 2 มีตัวแทนแล้ว) และเมื่อทุกออบเจ็คถูกครอบคลุมก็จะเป็นอันสิ้นสุดการทำงาน 1 รอบ สรุปในการทำงานรอบแรกจะได้โปรโตไทป์ผลลัพธ์ คือ {A1, A4, B2}

จากนั้นจะเริ่มทำงานในรอบที่สอง โดยขั้นตอนการทำงานจะเหมือนกับรอบแรก ยกเว้น Candidate Set ที่ในรอบแรกจะนั้นทำการเลือกโปรโตไทป์จากออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอน แต่เมื่อผ่านการทำงานรอบแรกที่ทำการกรองออบเจ็คที่ซ้ำซ้อนออกไปจนเหลือออบเจ็คที่น่าจะเป็นโปรโตไทป์เพียง 3 ออบเจ็ค คือ ออบเจ็ค A1, A4 และ B2 เท่านั้น ดังนั้น Candidate Set ของการทำงานรอบที่สองก็คือ {A1, A4, B2} และในกระบวนการเลือกออบเจ็คต่างชนิดก็จะต้องเลือกจากออบเจ็ค 3 ออบเจ็คนี้เท่านั้น ดังแสดงตัวอย่างในรูปที่ 3.9

รูปที่ 3.9 การเลือก NUN ในรอบที่สองของออบเจ็ค A1 จากชุดข้อมูลตัวอย่างที่ 2

เมื่อพิจารณาออบเจ็ค A3 ในรูปที่ 3.9 เทียบกับออบเจ็ค A3 ในตารางที่ 3.3 จะเห็นว่า NUN ของออบเจ็ค A3 เปลี่ยนไป จากเดิมที่เป็นออบเจ็ค B1 แต่เมื่อผ่านการทำงานรอบแรกแล้วจะเห็นว่าออบเจ็ค B1 ไม่ได้อยู่ใน Candidate Set ของการทำงานรอบที่สอง (เพราะไม่ได้ถูกเลือกเป็นโปรโตไทป์ในการทำงานรอบแรก) ดังนั้นเมื่อพิจารณาข้อมูลจากใกล้ไปไกลในกระบวนการเลือกออบเจ็คต่างชนิด (ตามที่อธิบายไปในหัวข้อ "3.1.2.2-การเลือกออบเจ็คต่างชนิด") จึงไม่เลือกออบเจ็ค B1 เป็น NUN ของออบเจ็ค A1 และเมื่อพิจารณาต่อไปก็จะเห็นว่าออบเจ็ค B2 อยู่ใน Candidate Set และมีชนิดต่างกับออบเจ็ค A3 ดังนั้นจึงทำการเลือกออบเจ็ค B2 เป็น NUN ของออบเจ็ค A3 ซึ่งจะเห็นว่ารัศมีขอบเขตต่างชนิดของออบเจ็ค A3 จะมีขนาดใหญ่ขึ้น ทำให้มีออบเจ็คที่สามารถครอบคลุมออบเจ็ค A3 ได้เพิ่มขึ้น จากเดิมที่มีแค่ออบเจ็ค A1, A3 และ A4 เท่านั้นที่ครอบคลุมออบเจ็ค A3 แต่เมื่อออบเจ็ค B1 ไม่ได้ถูกพิจารณาในกระบวนการเลือกออบเจ็คต่างชนิด จึงทำให้มีออบเจ็คที่สามารถครอบคลุมออบเจ็ค A3 ได้เพิ่มขึ้น ซึ่งก็คือออบเจ็ค A2 และเมื่อทำการเลือก NUN ในรอบที่สองครบทุกออบเจ็คแล้ว ก็จะสามารถแสดงได้ดังในตารางที่ 3.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.9 การเลือก NUN ในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2

Candidate Set = {A1, A4, B2}

A1	A1	A2	A3	B2	B1	A4	B3
A2	A2	A1	A3	B2	B1	B3	A4
A3	A3	A1	A4	B1	A2	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

จากนั้นก็จะพิจารณาความครอบคลุมของแต่ละออบเจ็ค โดยใส่ค่า "1" หรือ "0" (ตามที่อธิบายไปในหัวข้อ "3.1.2.2-การเลือกออบเจ็คต่างชนิด") และเมื่อพิจารณาความครอบคลุมครบทุกออบเจ็คแล้ว สามารถแสดงได้ดังในตารางที่ 3.10

ตารางที่ 3.10 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2

Candidate = {A1, A4, B2}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	1	0	0	0	0
A3	1	1	1	1	0	0	0
A4	0	0	1	1	0	0	0
B1	0	0	0	0	1	1	0
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1
CV	3	3	4	2	3	3	2

เมื่อพิจารณาตารางที่ 3.10 เทียบกับตารางที่ 3.5 จะเห็นว่าค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คเปลี่ยนไป ตัวอย่างที่เห็นได้ชัดเจน คือ ออบเจ็ค A3 จากตารางที่ 3.5 (ซึ่งเป็นการทำงานรอบแรก) จะเห็นว่ามีค่าความสามารถในการครอบคลุมเท่ากับ "3" แต่ในตารางที่ 3.10 (ซึ่งเป็นการทำงานรอบที่สอง) มีค่าความสามารถในการครอบคลุมเท่ากับ "4" เพราะออบเจ็ค B1 ไม่ได้ถูกพิจารณาในกระบวนการเลือกออบเจ็คต่างชนิดในการทำงานรอบที่สอง (เพราะออบเจ็ค B1 ไม่ได้อยู่ใน Candidate Set ของการทำงานรอบที่สอง) เป็นผลให้ออบเจ็ค A3 สามารถครอบคลุมออบเจ็ค A2 ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แต่อย่างไรก็ตามไม่สามารถเลือกออบเจ็ค A3 เป็นโปรโตไทป์ได้ เพราะออบเจ็ค A3 ไม่ได้อยู่ใน Candidate Set ของการทำงานรอบที่สอง ทั้งๆ ที่ถ้าเลือกออบเจ็ค A3 เพียงออบเจ็คเดียวจะสามารถครอบคลุมออบเจ็ค A1, A2, A3 และ A4 ได้ทั้งหมดก็ตาม ซึ่งนี่เองเป็นหนึ่งในข้อด้อยของวิธี MCS

ต่อมาในขั้นตอนกระบวนการการเลือกออบเจ็คในการทำงานรอบที่สองก็จะเหมือนกับกระบวนการเลือกออบเจ็คในรอบแรก กล่าวคือจะทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจาก Candidate Set เป็นโปรโตไทป์ ซึ่งในที่นี้ก็คือออบเจ็ค A1 หรือออบเจ็ค B2 เพราะทั้ง 2 ออบเจ็คอยู่ใน Candidate Set และมีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ "3" เท่ากัน โดยในที่นี้สมมุติว่าทำการเลือกออบเจ็ค A1 เป็นโปรโตไทป์ จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค A1, A2 และ A3 ที่ถูกครอบคลุมโดยออบเจ็ค A1 ที่ถูกเลือกเป็นโปรโตไทป์ไปแล้วออก ดังแสดงในตารางที่ 3.11

ตารางที่ 3.11 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 เป็นโปรโตไทป์แล้ว

Candidate Set = {A1, A4, B2}

Prototype Set = {A1}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	X	0	0	0	0
A3	X	X	X	X	0	0	0
A4	0	0	X	1	0	0	0
B1	0	0	0	0	1	1	0
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1
CV	0	0	0	1	3	3	2

จากนั้นทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจาก Candidate Set ลำดับต่อไปเป็นโปรโตไทป์ ซึ่งจากตารางที่ 3.11 จะเห็นว่าออบเจ็ค B2 อยู่ใน Candidate Set และมีค่าความสามารถในการครอบคลุมมากที่สุดเท่ากับ "3" ดังนั้นจึงทำการเลือกออบเจ็ค B2 เป็นโปรโตไทป์ จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค B1, B2 และ B3 ที่ถูกครอบคลุมโดยออบเจ็ค B2 ที่ถูกเลือกเป็นโปรโตไทป์ไปแล้ว ดังแสดงในตารางที่ 3.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.12 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1 และ B2 เป็นโปรโตไทป์แล้ว

Candidate Set = {A1, A4, B2}

Prototype Set = {A1, B2}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	X	0	0	0	0
A3	X	X	X	X	0	0	0
A4	0	0	X	1	0	0	0
B1	0	0	0	0	X	X	0
B2	0	0	0	0	X	X	X
B3	0	0	0	0	X	X	X
CV	0	0	0	1	0	0	0

ขั้นตอนสุดท้ายก็จะทำการเลือกออบเจ็ค A4 ที่เหลืออยู่เพียงตัวเดียวเป็นโปรโตไทป์ และเมื่อทำการเลือกออบเจ็ค A4 เป็นโปรโตไทป์แล้ว ค่าความสามารถในการครอบคลุมของออบเจ็ค A4 ก็จะถูกลดทอนออก ดังแสดงในตารางที่ 3.13

ตารางที่ 3.13 ค่าความสามารถในการครอบคลุมในการทำงานรอบที่สองของชุดข้อมูลตัวอย่างที่ 2 หลังเลือกออบเจ็ค A1, B2 และ A4 เป็นโปรโตไทป์แล้ว

Candidate Set = {A1, A4, B2}

Prototype Set = {A1, A4, B2}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	X	0	0	0	0
A3	X	X	X	X	0	0	0
A4	0	0	X	X	0	0	0
B1	0	0	0	0	X	X	0
B2	0	0	0	0	X	X	X
B3	0	0	0	0	X	X	X
CV	0	0	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อพิจารณาตารางที่ 3.13 จะเห็นว่าค่าความสามารถในการครอบคลุมของทุกออบเจ็คเท่ากับ "0" เพราะทุกออบเจ็คในชุดข้อมูลที่ 2 ถูกครอบคลุมโดยโปรโตไทป์ที่เลือกแล้ว และเมื่อทุกออบเจ็คถูกครอบคลุมหมดแล้วก็เป็นอันสิ้นสุดการทำงานในรอบที่สอง สรุปในการทำงานรอบที่สองจะได้โปรโตไทป์ผลลัพธ์ คือ {A1, A4, B̄2}

ซึ่งถ้าจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ในรอบที่สองนี้มีจำนวนน้อยกว่าจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ในรอบแรกก็จะทำงานในรอบที่สามต่อไป แต่ถ้าจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ในรอบที่สองเท่ากับจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ในรอบแรก ก็จะเป็นอันสิ้นสุดการทำงานของวิธี MCS

จากตัวอย่างข้างต้นก็จะเห็นว่าจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ในรอบที่สองมีเท่ากับจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ในรอบแรก ดังนั้นจึงเป็นอันสิ้นสุดการทำงานของวิธี MCS ดังนั้นสรุปได้ว่าวิธี MCS สามารถลดจำนวนโปรโตไทป์ผลลัพธ์ของชุดข้อมูลตัวอย่างที่ 2 จากเดิม 7 ตัวให้เหลือเพียง 3 ตัว (หมายความว่าสามารถใช้ข้อมูลเพียง 3 ตัวแทนข้อมูลทั้ง 7 ตัว)

จากผลลัพธ์การทำงานของวิธี MCS ที่ได้แสดงไปข้างต้นจะเห็นว่าการเริ่มต้นทำงานโดยใช้ออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนในการคำนวณความครอบคลุมให้ผลลัพธ์ที่ดี แต่เมื่อทดลองเริ่มต้นทำงาน โดยใช้เพียงบางออบเจ็คในชุดข้อมูลฝึกสอนเท่านั้นในการคำนวณความครอบคลุม ปรากฏว่าสามารถให้ผลลัพธ์ที่ดีกว่า ตัวอย่างเช่น กำหนด Candidate Set ตัวอย่างที่ 1 คือ {A2, B3} เท่านั้น สามารถแสดงขั้นตอนการทำงานได้ดังนี้

ขั้นตอนการทำงานก็จะเริ่มต้นด้วยกระบวนการเลือกออบเจ็คต่างชนิด (ตามที่ได้อธิบายในหัวข้อ "3.1.2.2-การเลือกออบเจ็คต่างชนิด") ดังแสดงผลลัพธ์ในตารางที่ 3.14

ตารางที่ 3.14 การเลือก NUN ในการทำงานของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1

Candidate Set = {A2, B3}

A1	A1	A2	A3	B2	B1	A4	B3
A2	A2	A1	A3	B2	B1	B3	A4
A3	A3	A1	A4	B1	A2	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนต่อมาจะพิจารณาความครอบคลุมของแต่ละออบเจ็ค โดยใส่ค่า “1” หรือ “0” (ตามที่ได้อธิบายในหัวข้อ “3.1.2.2-การพิจารณาความครอบคลุม”) และทำการคำนวณค่าความสามารถในการครอบคลุม ดังแสดงผลลัพธ์ในตารางที่ 3.15

ตารางที่ 3.15 ค่าความสามารถในการครอบคลุมของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1

Candidate Set = {A2, B3}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	1	0	0	0
A2	1	1	1	0	0	0	0
A3	1	1	1	1	0	0	0
A4	1	1	1	1	0	0	0
B1	0	0	0	0	1	1	1
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1
CV	4	4	4	3	3	3	3

ขั้นตอนต่อมาก็คือกระบวนการเลือกออบเจ็ค ซึ่งหลักการเลือกก็เหมือนเดิมกล่าวคือทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจาก Candidate Set เป็นโปรโตไทป์ โดยในที่นี้ก็คือออบเจ็ค A2 เพราะออบเจ็ค A2 อยู่ใน Candidate Set และมีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ “4” จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค A1, A2, A3 และ A4 ที่ถูกครอบคลุมโดยออบเจ็ค A2 แล้วออก ดังแสดงผลลัพธ์ในตารางที่ 3.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.16 ค่าความสามารถในการครอบคลุมของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1 หลังเลือกออบเจ็ค A2 เป็นโปรโตไทป์แล้ว

Candidate Set = {A2, B3}

Prototype Set = {A2}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	X	0	0	0
A2	X	X	X	0	0	0	0
A3	X	X	X	X	0	0	0
A4	X	X	X	X	0	0	0
B1	0	0	0	0	1	1	1
B2	0	0	0	0	1	1	1
B3	0	0	0	0	1	1	1
CV	0	0	0	0	3	3	3

ขั้นตอนสุดท้ายก็จะทำการเลือกออบเจ็ค B3 ที่เหลืออยู่เพียงตัวเดียวเป็นโปรโตไทป์ และเมื่อทำการเลือกออบเจ็ค B3 เป็นโปรโตไทป์แล้ว ค่าความสามารถในการครอบคลุมของออบเจ็ค B1, B2 และ B3 ที่ถูกครอบคลุมโดยออบเจ็ค B3 แล้วออกก็จะถูกลดทอนออก ดังแสดงในตารางที่ 3.17

ตารางที่ 3.17 ค่าความสามารถในการครอบคลุมของชุดข้อมูลตัวอย่างที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 1 หลังเลือกออบเจ็ค A2 และ B3 เป็นโปรโตไทป์แล้ว

Candidate Set = {A2, B3}

Prototype Set = {A2, B3}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	X	0	0	0
A2	X	X	X	0	0	0	0
A3	X	X	X	X	0	0	0
A4	X	X	X	X	0	0	0
B1	0	0	0	0	X	X	X
B2	0	0	0	0	X	X	X
B3	0	0	0	0	X	X	X
CV	0	0	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อพิจารณาตารางที่ 3.17 จะเห็นว่าค่าความสามารถในการครอบคลุมของทุกออบเจ็คเท่ากับ "0" เพราะทุกออบเจ็คในชุดข้อมูลที่ 2 ถูกครอบคลุมโดยโปรโตไทป์ที่เลือกแล้ว และเมื่อทุกออบเจ็คถูกครอบคลุมหมดแล้วก็เป็นอันสิ้นสุดการทำงาน 1 รอบ สรุปในการทำงานรอบแรกจะได้โปรโตไทป์ผลลัพธ์ คือ {A2, B3}

ถึงแม้ว่ายังต้องทำงานในรอบที่สองต่อไปอีก แต่จะเห็นได้ชัดเจนว่าโปรโตไทป์ผลลัพธ์ที่ได้ก็จะเหมือนเดิมก็คือ {A2, B3} ซึ่งจะเห็นว่ามีจำนวนโปรโตไทป์ผลลัพธ์มีน้อยกว่าเดิม และเมื่อเปรียบเทียบกับการกำหนด Candidate Set แบบเดิม (ตารางที่ 3.5) จะเห็นได้ชัดเจนว่าไม่มีทางที่จะทำการเลือกออบเจ็ค A2 และออบเจ็ค B3 เป็นโปรโตไทป์ได้เลย เพราะทั้ง 2 ออบเจ็คไม่ได้มีค่าความสามารถในการครอบคลุมสูงสุด ซึ่งนี่เป็นเครื่องยืนยันได้อย่างดีว่าการเริ่มต้นการทำงานของวิธี MCS โดยใช้ออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนเป็น Candidate Set ซึ่งจะนำใช้ในการคำนวณความครอบคลุมอีกทีนั้น ไม่ใช่วิธีที่ดีที่สุด

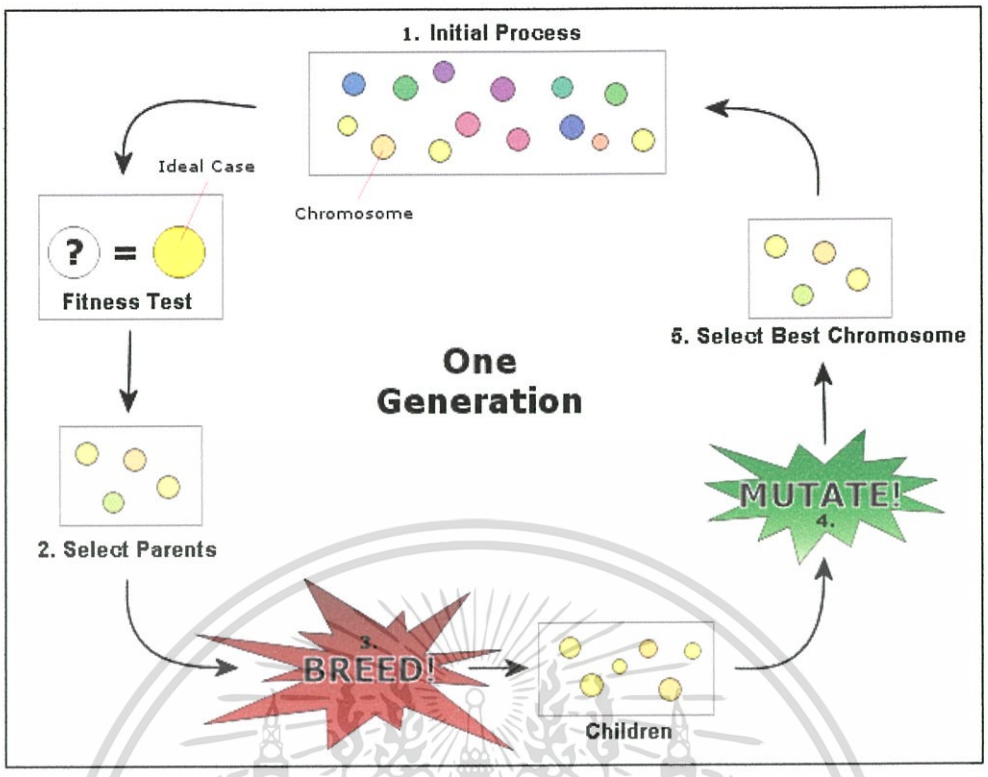
จึงเป็นที่มาของปัญหาในวิทยานิพนธ์ฉบับนี้ที่จะต้องหาวิธีกำหนดออบเจ็คที่ใช้ในการเริ่มต้นการทำงานของวิธี MCS ที่จะทำให้ได้โปรโตไทป์ผลลัพธ์ที่ดีขึ้น

3.2 Genetic Algorithms (GAs)

GAs [14] เป็นวิธีการที่เชิงวิทยาศาสตร์ที่ใช้ในการหาผลลัพธ์โดยการประมาณค่าเพื่อใช้ในการหาคำตอบที่ดีที่สุด (Optimization) ถูกคิดขึ้นโดย John Holland โดยมีหลักการทำงานพื้นฐานที่เลียนแบบมาจากระบบวิวัฒนาการทางธรรมชาติสำหรับพันธุกรรมของสิ่งมีชีวิต หรือที่เรียกกันว่า "ทฤษฎีวิวัฒนาการทางพันธุกรรม" ของ Charles Darwin หรือ "Darwinian Theory" [15] ที่เป็นทฤษฎีที่อธิบายเกี่ยวกับวัฏจักรของพันธุกรรมว่าเกิดจากความหลากหลายทางพันธุกรรม (Genetic Diversity) ที่ได้รับการคัดสรรทางธรรมชาติ (Natural Selection) ซึ่งโดยธรรมชาติแล้ว สิ่งมีชีวิตที่สามารถปรับตัวให้เข้ากับสภาพแวดล้อมที่เป็นอยู่ได้ ก็จะมีโอกาสอยู่รอดมากกว่าสิ่งมีชีวิตที่ไม่สามารถปรับตัวได้ หรือที่เรียกว่า "The Survival of the Fittest" และเมื่อสิ่งมีชีวิตที่มีความเหมาะสมกับสภาพแวดล้อมมาผสมพันธุ์กันก็จะได้ลูกหลานออกมา ซึ่งลูกหลานที่ได้ก็จะได้รับการถ่ายทอดวิวัฒนาการที่มีความเหมาะสมกับสภาพแวดล้อมด้วย หรือพูดง่ายๆ ก็คือ "ลูกไม่หล่นไม่ไกลต้น"

GAs เป็นการค้นหาคำตอบแบบกว้าง (Global Search) ที่อาศัยกระบวนการทางพันธุกรรม (Genetic Process) ในการค้นหาคำตอบ ซึ่งสามารถนำไปประยุกต์ใช้ในการแก้ปัญหาต่างๆ ได้มากมาย ซึ่งการนำไปแก้ปัญหาแต่ละแบบก็จะต้องประยุกต์ใช้ GAs ในรูปแบบที่แตกต่างกันออกไป แต่ขั้นตอนการทำงานหลักๆ จะเหมือนกัน ดังแสดงในรูปที่ 3.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 แผนภาพวัฏจักรการทำงานของ GAs

จากรูปที่ 3.10 แสดงแผนภาพวัฏจักรการทำงานของ GAs ซึ่งจะเห็นว่าขั้นตอนการทำงานเริ่มต้นด้วยการแทนปัญหาที่ต้องการหาคำตอบให้อยู่ในรูปของโครโมโซมก่อน จากนั้นทำการสุ่มสร้างโครโมโซมเพื่อใช้เป็นประชากรเริ่มต้น โดยจำนวนโครโมโซมจะเท่ากับจำนวนประชากรตามที่กำหนดไว้ล่วงหน้า จากนั้นจะนำโครโมโซมในประชากรเริ่มต้นทั้งหมดไปเทียบกับคำตอบที่ต้องการโดยใช้ฟังก์ชั่นคำนวณค่าความเหมาะสม (Fitness Function) เพื่อวัดค่าความเหมาะสมของแต่ละโครโมโซม โดยทำการเทียบกับค่าของคำตอบที่ต้องการ โดยฟังก์ชั่นคำนวณความเหมาะสมจะให้ผลลัพธ์เป็นค่าความเหมาะสม (Fitness Value) ต่อมาจะทำการสุ่มเลือกโครโมโซมขึ้นมาเป็นพ่อ-แม่พันธุ์ โดยอัตราส่วนความน่าจะเป็นในการสุ่มนั้นจะแปรผันตรงกับค่าความเหมาะสม ของแต่ละโครโมโซม กล่าวคือโครโมโซมที่มีความเหมาะสมมากก็จะมีโอกาสถูกเลือกมาเป็นพ่อ-แม่พันธุ์ได้มาก จากนั้นจะนำโครโมโซมพ่อ-แม่มาผสมกัน (Crossover) เพื่อให้ได้โครโมโซมลูกออกมา (ซึ่งก็ตรงตามหลักการ “ลูกไม้หล่นไม่ไกลต้น” เพราะโครโมโซมพ่อ-แม่ที่มีความเหมาะสมมากก็น่าจะให้โครโมโซมลูกที่มีค่าความเหมาะสมมากเช่นกัน) ต่อมาโครโมโซมลูกที่ได้นั้นก็อาจจะเกิดการกลายพันธุ์ (Mutation) ขึ้นได้ และสุดท้ายทำการคัดสรรโครโมโซมที่มีค่าความเหมาะสมมากที่สุดจากโครโมโซมพ่อ-แม่-ลูก ไปเป็นประชากรในรอบวิวัฒนาการต่อไป (ซึ่งก็ตรงตามหลักการ “The Survival of the Fittest” ที่โครโมโซมที่มีความเหมาะสมมากกว่าก็จะมีโอกาสอยู่รอดมากกว่า)

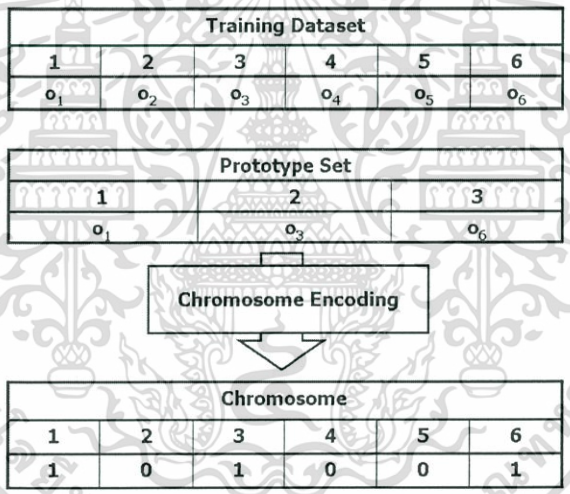
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1 นิยามที่ควรทราบเกี่ยวกับ GAs

3.2.1.1 โครโมโซม (Chromosome)

โครโมโซม คือ รูปแบบการแทนปัญหาที่ต้องการหาผลลัพธ์โดยใช้ GAs โดยผ่านการเข้ารหัสโครโมโซม (Chromosome Encoding) ซึ่งรูปแบบการแทนก็จะแตกต่างกันไป ขึ้นกับว่าจะนำ GAs ไปใช้แก้ปัญหาอะไร

โดยการประยุกต์ใช้ GAs มาใช้ในการเลือกโปรโตไทป์ในในวิทยานิพนธ์ฉบับนี้จะอ้างอิงจากตามที่ได้เสนอไว้ในบทความที่มีชื่อว่า "Nearest Prototype Classifier: Clustering, Genetic Algorithms, or Random Search?" [5] โดย L.I. Kuncheva และ J.C. Bezdek ที่ทำการแทนโปรโตไทป์ให้อยู่ในรูปของ "Binary String" ที่มีความยาวเท่ากับจำนวนออบเจ็คในชุดข้อมูลฝึกสอนโดยมีข้อกำหนดว่าหากเลือกออบเจ็คตำแหน่งใดในชุดข้อมูลฝึกสอนไปเป็นโปรโตไทป์ก็กำหนดให้ค่าบิตในตำแหน่งนั้นมีค่าเป็น "1" ส่วนตำแหน่งอื่นๆที่ออบเจ็คไม่ได้ถูกเลือกไปเป็นโปรโตไทป์กำหนดให้มีค่าเป็น "0" ดังแสดงตัวอย่างในรูปที่ 3.11



รูปที่ 3.11 ตัวอย่างการเข้ารหัสโครโมโซม

จากรูปที่ 3.11 แสดงตัวอย่างการเข้ารหัสโครโมโซม โดยตาราง "Training Dataset" แสดงถึงชุดข้อมูลฝึกสอนที่มีข้อมูลจำนวน 6 ตัว ซึ่งประกอบไปด้วย O_1, O_2, O_3, O_4, O_5 และ O_6 ซึ่งเรียงอยู่ในตำแหน่งที่ 1 ไปจนถึงตำแหน่งที่ 6 ตามลำดับ ส่วนตาราง "Prototype Set" แสดงถึงเซตของออบเจ็คที่เลือกเป็นโปรโตไทป์ ที่มีสมาชิกอยู่ทั้งหมด 3 ตัว ซึ่งประกอบไปด้วย O_1, O_3 และ O_6 โปรโตไทป์ที่เลือกได้เหล่านี้สามารถเข้ารหัสโครโมโซมให้อยู่ในรูปแบบ "Binary String" ที่มีความยาวจำนวน 6 บิต ซึ่งเท่ากับจำนวนสมาชิกของชุดข้อมูลฝึกสอนโดยที่บิตสตริงมีบิตในตำแหน่งที่ 1 , 2 และ 6 มีค่าเป็น "1" เพราะออบเจ็คที่อยู่ในตำแหน่งที่ 1, 2 และ 6 ในชุดข้อมูลถูกเลือกไปเป็นโปรโตไทป์ ส่วนตำแหน่งอื่นมีค่าเป็น "0"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1.2 ประชากร (Population)

ประชากร คือ กลุ่มของโครโมโซมที่จะนำไปใช้เป็นพ่อ-แม่พันธุ์โดยกระบวนทางทางพันธุกรรมต่อไป โดยจะมีจำนวนตามที่ได้กำหนดไว้ล่วงหน้า

3.2.1.3 จำนวนประชากร (N_{pop})

จำนวนประชากร คือ จำนวนโครโมโซมทั้งหมดที่อยู่ในประชากร โดยจะมีจำนวนตามที่กำหนดไว้ล่วงหน้า

3.2.1.4 จำนวนเจเนอเรชั่น (N_{gen})

จำนวนเจเนอเรชั่น คือ จำนวนรอบการทำงานของ GAs โดยจะมีค่าตามที่กำหนดไว้ล่วงหน้า

3.2.1.5 ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini})

ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น คือ ค่าความน่าจะเป็นในการกำหนดค่าแต่ละบิตในโครโมซมให้มีค่าเป็น "1" โดยจะมีค่าตามที่ได้กำหนดไว้ล่วงหน้า

3.2.1.6 ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c)

ค่าความน่าจะเป็นในการครอสโอเวอร์ คือ ค่าความน่าจะเป็นในการกำหนดว่าแต่ละบิตในโครโมโซมจะเกิดการครอสโอเวอร์หรือไม่ โดยจะมีค่าตามที่ได้กำหนดไว้ล่วงหน้า

3.2.1.7 ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m)

ค่าความน่าจะเป็นในการกลายพันธุ์ คือ ค่าความน่าจะเป็นในการกำหนดว่าแต่ละบิตใครโครโมโซมจะเกิดการกลายพันธุ์หรือไม่ โดยจะมีค่าตามที่ได้กำหนดไว้ล่วงหน้า

3.2.1.8 ฟังก์ชั่นคำนวณค่าความเหมาะสม (Fitness Function)

ฟังก์ชั่นคำนวณค่าความเหมาะสมเป็นฟังก์ชั่นที่ใช้วัดความเหมาะสมของโครโมโซมเทียบกับเกณฑ์ หรือคำตอบที่เราต้องการ ซึ่งฟังก์ชั่นคำนวณความเหมาะสมที่ใช้ก็จะแตกต่างกันไปขึ้นอยู่ว่านำเอา GAs ไปใช้แก้ปัญหาอะไร ส่วนในวิทยานิพนธ์ฉบับนี้ใช้ฟังก์ชั่นคำนวณความเหมาะสมตามบทความที่มีชื่อว่า "Nearest Prototype Classifier: Clustering, Genetic Algorithms, or Random Search?" ซึ่งสามารถแสดงเป็นสมการคณิตศาสตร์ได้ดังนี้

$$F(S) = \frac{Accuracy(S) - \alpha * Size(S)}{N_TrainingDataset} \tag{3.1}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยกำหนดให้

- $\mathbb{F}(S)$ คือ ฟังก์ชั่นคำนวณค่าความเหมาะสมของโครโมโซม S
- S คือ โครโมโซมที่ถูกเข้ารหัสแทนโปรโตไทป์
- $Accuracy(S)$ คือ ฟังก์ชันแสดงจำนวนออบเจ็คในชุดข้อมูลฝึกสอนที่ถูกจำแนกโดยใช้ NN Rule ที่มีโปรโตไทป์ S เป็นข้อมูลอ้างอิงได้ถูกต้อง
- $Size(S)$ คือ ฟังก์ชั่นแสดงจำนวนออบเจ็คที่ถูกเลือกเป็นโปรโตไทป์ของโปรโตไทป์ S
- $N_TrainingDataset$ คือ จำนวนออบเจ็คทั้งหมดของชุดข้อมูลฝึกสอน
- α คือ ค่าสัมประสิทธิ์ของการถ่วงน้ำหนัก (Weighted Coefficient) เพื่อถ่วงความสำคัญของพจน์ $Size(S)$ กับพจน์ $Accuracy(S)$ (โดยในวิทยานิพนธ์ฉบับนี้กำหนดให้ α มีค่าเท่ากับ $1/N_TrainingDataset$ เพราะให้ความสำคัญกับ $Accuracy(S)$ มากกว่า)

ค่าความเหมาะสมจะมีค่าอยู่ในช่วงตั้งแต่ "-1" ถึง "1" โดยหากค่าของฟังก์ชั่น $Accuracy(S)$ มีค่ามาก นั่นคือการจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ S เป็นข้อมูลอ้างอิงสามารถจำแนกออบเจ็คในชุดข้อมูลฝึกสอนได้เป็นจำนวนมาก และค่า $Size(S)$ มีค่าน้อย นั่นคือจำนวนออบเจ็คที่ถูกเลือกเป็นโปรโตไทป์ของโปรโตไทป์ S มีจำนวนน้อย ก็จะทำให้ค่าความเหมาะสม มีค่ามาก (ใกล้ 1) หมายความว่าโครโมโซม S เป็นโครโมโซมที่ดี ควรได้รับการพิจารณานำไปใช้เป็นพ่อ-แม่พันธุ์ต่อไป

ในทางกลับกันหากค่าของฟังก์ชั่น $Accuracy(S)$ มีค่าน้อย นั่นคือการจำแนกโดยใช้ NN Rule ที่มีโปรโตไทป์ S เป็นข้อมูลอ้างอิง สามารถจำแนกออบเจ็คในชุดข้อมูลฝึกสอนได้จำนวนน้อย และค่า $Size(S)$ มีค่ามาก นั่นคือจำนวนออบเจ็คที่ถูกเลือกเป็นโปรโตไทป์ของโปรโตไทป์ S มีเป็นจำนวนมาก ก็จะทำให้ ค่าความเหมาะสมมีค่าน้อย (ใกล้ -1) หมายความว่าโครโมโซม S เป็นโครโมโซมที่ไม่ดี ไม่ควรได้รับการพิจารณานำไปใช้เป็นพ่อ–แม่พันธุ์

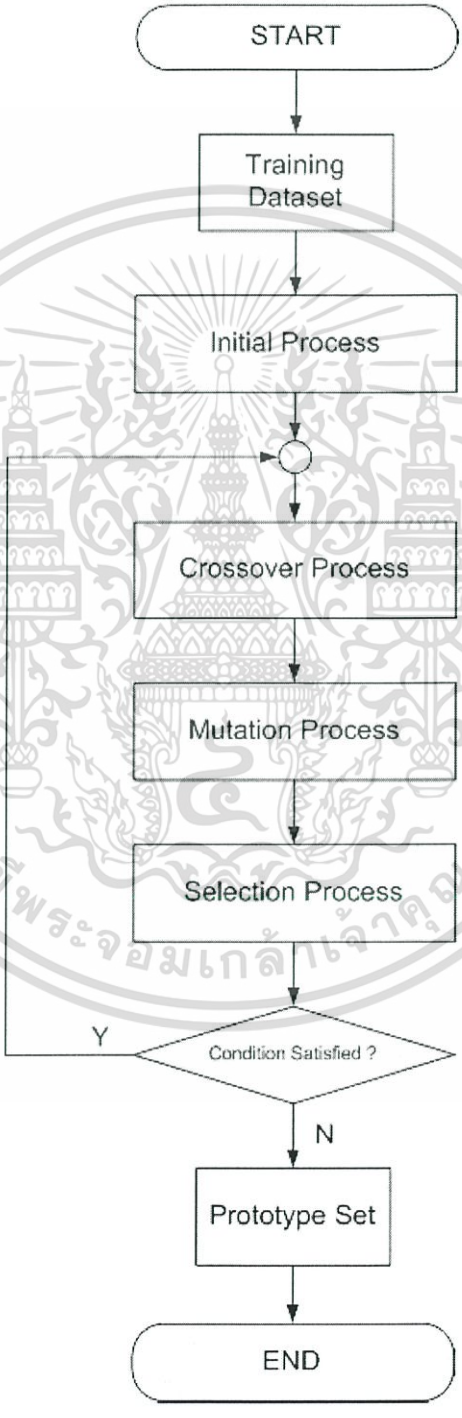
3.2.1.9 โครโมโซมที่ดีที่สุด (The Best Chromosome)

โครโมโซมที่ดีที่สุดเป็นโครโมโซมที่ดีที่สุดเท่าที่เคยหาผลลัพธ์ผ่านมา โดยอาจจะเป็นโครโมโซมที่มีค่าความเหมาะสมสูงสุด หรือเป็นโครโมโซมที่ตรงตามเงื่อนไขที่กำหนดไว้ก็ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 รายละเอียดการทำงานของ GAs

ตามที่ได้กล่าวไปข้างต้นว่า GAs นั้นมีหลักการทำงานพื้นฐานที่เลียนแบบมาจากระบบวิวัฒนาการทางธรรมชาติสำหรับพันธุกรรมของสิ่งมีชีวิต ดังนั้นกระบวนการทำงานต่างๆ ก็จะเลียนแบบมาจากระบบวิวัฒนาการทางธรรมชาติเช่นกัน โดยเรียกรวมๆ ว่า "กระบวนการทางพันธุกรรม (Genetic Process)" ดังแสดงแผนภาพขั้นตอนการทำงานของ GAs ในรูปที่ 3.12



รูปที่ 3.12 แผนภาพขั้นตอนการทำงานของ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.12 แสดงแผนภาพขั้นตอนการทำงานของ GAs โดยเริ่มต้นการทำงานด้วยกระบวนการสุ่มกำหนดค่าเริ่มต้น (Initial Process) ที่ทำการสุ่มสร้างโครโมโซมขึ้นตามจำนวนที่กำหนดไว้ล่วงหน้า ต่อมากระบวนการครอสโอเวอร์ (Crossover Process) ที่ทำการสุ่มเลือกโครโมโซมพ่อ-แม่เพื่อนำมาผสมกันให้ได้โครโมโซมลูก จากนั้นกระบวนการกลายพันธุ์ (Mutation Process) ที่ทำการสุ่มเปลี่ยนแปลงค่าภายในโครโมโซมลูก สุดท้ายกระบวนการคัดสรรประชากรรุ่นถัดไป (Selection Process) ที่ทำการเลือกเก็บโครโมโซมบางตัวเพื่อนำไปเป็นประชากรในรุ่นถัดไป หลังจากนั้นก็จะวนทำงานซ้ำไปเรื่อยๆ จนกว่าจะครบตามเงื่อนไขที่กำหนดไว้ล่วงหน้า ซึ่งโดยปกติแล้ว GA จะมีเงื่อนไขการสิ้นสุดการทำงานได้ 2 กรณี ดังนี้ กรณีแรก คือ เมื่อเฟ้นสุ่มหาคำตอบได้ตามเกณฑ์ตามที่กำหนดไว้ล่วงหน้า ส่วนกรณีที่สอง คือ เมื่อวนทำงานไปจนครบตามจำนวนเจเนอเรชั่นหรือตามเวลาที่กำหนดไว้ล่วงหน้า โดยในวิทยานิพนธ์ฉบับนี้กำหนดให้ GAs ทำงานไปจนครบรอบตามเวลาที่กำหนดไว้ล่วงหน้า เพราะเป็นการเฟ้นสุ่มเพื่อหาโปรโตไทป์ที่มีจำนวนน้อยที่สุด จึงไม่สามารถกำหนดเกณฑ์คำตอบที่ดีได้

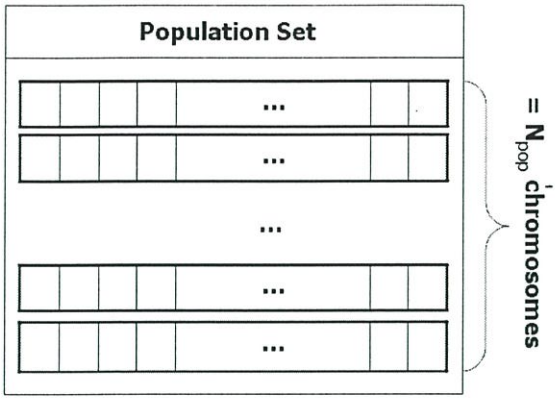
3.2.2.1 กระบวนการกำหนดค่าเริ่มต้น (Initial Process)

กระบวนการกำหนดค่าเริ่มต้นเป็นกระบวนการที่ทำงานแรกสุด และกระบวนการนี้จะทำงานเพียงครั้งเดียวในตอนแรกเท่านั้น โดยจะทำการเข้ารหัสโครโมโซมเพื่อแทนชุดข้อมูลฝึกสอนให้อยู่ในรูปโครโมโซม จากนั้นจะทำการสร้างโครโมโซมจำนวนประชากรตามที่กำหนดไว้ล่วงหน้า และจะเรียกกลุ่มของโครโมโซมเหล่านี้ว่าเป็น “ประชากร” เริ่มต้น จากนั้นก็จะทำการสุ่มกำหนดค่าเริ่มต้นให้กับประชากรเหล่านี้ ด้วยค่าความน่าจะเป็น P_{ini} ตามที่ได้กำหนดไว้ล่วงหน้า โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

- **การสร้างประชากรเริ่มต้น (Population Creation)**

การสร้างประชากรเริ่มต้นเป็นการที่จะสร้างโครโมโซมเพื่อใช้เป็นประชากรเริ่มต้น โดยจะมีจำนวนเท่ากับจำนวนประชากรตามที่กำหนดไว้ล่วงหน้า และแต่ละโครโมโซมมีขนาดเท่ากับจำนวนข้อมูลในชุดข้อมูลฝึกสอน (ตามที่อธิบายไปในหัวข้อที่ 3.2.2.1) ดังแสดงตัวอย่างในรูปที่ 3.13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

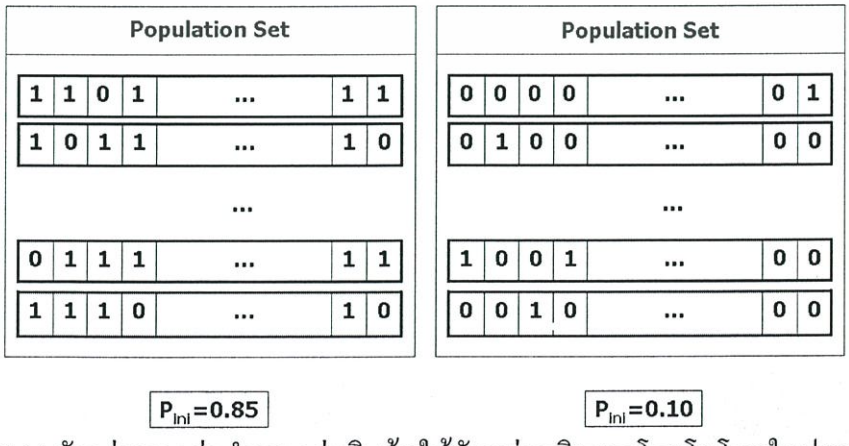


รูปที่ 3.13 การสร้างประชากรเริ่มต้น

- **การสุ่มกำหนดค่าให้กับประชากร (Population Initialization)**

การสุ่มกำหนดค่าให้กับประชากรเป็นการที่จะสุ่มกำหนดค่าให้กับแต่ละบิตของโครโมโซมในประชากร ด้วยความน่าจะเป็น P_{ini} ตามที่ได้กำหนดไว้ล่วงหน้า นั่นคือแต่ละบิตของโครโมโซมมีโอกาสที่จะสุ่มให้มีค่าเป็น "1" เท่ากับ P_{ini} และมีโอกาสที่จะสุ่มให้มีค่าเป็น "0" เท่ากับ $1-P_{ini}$

ด้วยค่า P_{ini} นี้ ทำให้สามารถควบคุมลักษณะของประชากรเริ่มต้นได้ เช่น ถ้าต้องการให้ประชากรโครโมโซมเริ่มต้นที่มีลักษณะหนาแน่น (แต่ละบิตมีโอกาสที่จะเป็น "1" ได้สูง) ก็กำหนดให้ $P_{ini} = 0.85$ (กำหนดให้ค่า P_{ini} ใกล้ "1") แต่ถ้าหากว่าต้องการให้ประชากรโครโมโซมเริ่มต้นมีลักษณะเบาบาง (แต่ละบิตมีโอกาสที่จะเป็น "1" ได้ต่ำ) ก็กำหนดให้ $P_{ini} = 0.10$ (กำหนดให้ค่า P_{ini} ใกล้ "0") เป็นต้น ซึ่งหากประชากรโครโมโซมเริ่มต้นหนาแน่น ค่าความถูกต้องในการจำแนกข้อมูล $Accuracy(S)$ ก็จะมีค่าสูง แต่จำนวนโปรโตไทป์ที่ถูกเลือก $Size(S)$ ก็จะมีจำนวนมากตามไปด้วย ในทางตรงกันข้ามหากประชากรโครโมโซมเริ่มต้นเบาบาง ค่าความถูกต้องในการจำแนกข้อมูล $Accuracy(S)$ ก็จะมีค่าต่ำ แต่จำนวนโปรโตไทป์ที่ถูกเลือก $Size(S)$ ก็จะมีจำนวนน้อยตามไปด้วย ดังแสดงตัวอย่างในรูปที่ 3.14



รูปที่ 3.14 ตัวอย่างการสุ่มกำหนดค่าเริ่มต้นให้กับแต่ละบิตของโครโมโซมในประชากร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากในรูปที่ 3.14 แสดงตัวอย่างสุ่มกำหนดค่าเริ่มต้นให้กับแต่ละบิตของโครโมโซมในประชากร โดยในด้านซ้ายของรูปที่ 3.13 แสดงประชากรที่ถูกสุ่มค่าเริ่มต้นโดยใช้ความน่าจะเป็น P_{ini} = 0.85 ซึ่งจะสังเกตเห็นได้ว่าโครโมโซมจะมีบิตที่มีค่าเป็น "1" อยู่อย่างหนาแน่น ส่วนในด้านขวาของรูปที่ 3.13 แสดงประชากรที่ถูกสุ่มค่าเริ่มต้นโดยใช้ความน่าจะเป็น P_{ini} = 0.10 ซึ่งจะสังเกตเห็นได้ว่าโครโมโซมจะมีบิตที่มีค่าเป็น "1" อยู่อย่างเบาบาง

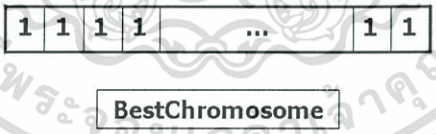
- **การสร้างโครโมโซมที่ดีที่สุด (The Best Chromosome Creation)**

การสร้างโครโมโซมที่ดีที่สุดเป็นการที่จะสร้างโครโมโซมที่ดีที่สุด โดยในวิทยานิพนธ์ฉบับนี้จะนิยามโครโมโซมที่ดีที่สุด คือ โครโมโซมที่เป็นไปตามเงื่อนไข 2 ข้อ ดังนี้

เงื่อนไขที่ 1 คือ โครโมโซมจะต้องมีความสอดคล้องกับชุดข้อมูลฝึกสอน

เงื่อนไขที่ 2 คือ โครโมโซมจะต้องมีค่า $Size(S)$ น้อยที่สุดในบรรดาโครโมโซมทั้งหมดที่ได้เคยมีการสร้างขึ้นมา

โดยในตอนเริ่มแรกโครโมโซมที่ดีที่สุดก็คือโครโมโซมที่มีค่าเป็น "1" ทุกบิต ซึ่งก็คือโครโมโซมที่มีออบเจ็คในชุดข้อมูลฝึกสอนทั้งหมดเป็นสมาชิก ซึ่งรับรองว่าโครโมโซมที่ดีที่สุดจะมีความสอดคล้องกับชุดข้อมูลฝึกสอนแน่นอน แต่มีโปรโตไทป์จำนวนมาก ซึ่งถ้าภายหลังถ้าเกิดมีโครโมโซมที่มีความสอดคล้องกับชุดข้อมูลฝึกสอนและมีจำนวนที่น้อยกว่ามากกว่าโครโมโซมที่ดีที่สุด โครโมโซมที่ดีที่สุดจะถูกปรับปรุงค่าให้เป็นโครโมโซมนั้นแทน (การปรับปรุงค่าโครโมโซมที่ดีที่สุดจะอยู่ในกระบวนการเลือกสรรประชากรรุ่นถัดไป) ดังแสดงตัวอย่างในรูปที่ 3.15



รูปที่ 3.15 การกำหนดค่าเริ่มต้นให้โครโมโซมที่ดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2.2 กระบวนการครอสโอเวอร์ (Crossover Process)

กระบวนการครอสโอเวอร์เป็นกระบวนการที่จะทำการสุ่มเลือกโครโมโซมจากประชากรมาใช้เป็นพ่อ-แม่พันธุ์เพื่อสร้างโครโมโซมลูกหลาน โดยการสุ่มจะใช้หลักการรูเล็ทวีล (Roulette Wheel Principle) ที่ว่าโอกาสที่โครโมโซมใดๆ จะถูกเลือกจะเป็นอัตราส่วนแปรผันตรงกับค่าความเหมาะสมของโครโมโซมนั้น กล่าวคือโครโมโซมที่มีค่าความเหมาะสมมากก็จะมีความน่าจะเป็นที่ถูกสุ่มเลือกได้สูง ในทางกลับกันโครโมโซมที่มีค่าความเหมาะสมน้อยก็จะมีความน่าจะเป็นที่ถูกสุ่มเลือกได้ต่ำ และเมื่อได้โครโมโซมพ่อ-แม่แล้ว ก็จะทำการนำโครโมโซมพ่อ-แม่มาทำการครอสโอเวอร์ ด้วยความน่าจะเป็น P_c ตามที่กำหนดไว้ล่วงหน้า โดยในวิทยานิพนธ์ฉบับนี้จะใช้ "Uniform Crossover" ซึ่งเป็นการสุ่มสลับค่าบิตระหว่างโครโมโซมพ่อ-แม่ ด้วยอัตราส่วนเท่าๆ กัน ดังนั้นจึงกำหนดให้ค่าความน่าจะเป็น $P_c = 0.5$ โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

- **การสร้างรูเล็ทวีล (Roulette Wheel Creation)**

การสร้างรูเล็ทวีลเป็นที่มีหลักการทำงานมาจาก "Roulette Wheel" ซึ่งเป็นอุปกรณ์ที่ใช้ในการพนันชนิดหนึ่ง มีลักษณะเป็นวงล้อหมุนได้ บนวงล้อมีช่องจำนวน 36 ช่องขนาดเท่ากัน เพื่อใช้บ่งบอกว่าความน่าจะเป็นที่จะหมุนตกช่องใดๆ นั้นมีโอกาสเท่ากันหมด จากหลักการนี้เองสามารถนำมาประยุกต์ใช้ได้โดยจากที่ได้อธิบายไปข้างต้นว่าโครโมโซมที่มีค่าความเหมาะสมมากก็จะมีความน่าจะเป็นที่ถูกสุ่มเลือกได้สูง ดังนั้นช่องสำหรับโครโมโซมก็มีขนาดกว้างตามไปด้วย ซึ่งทำให้โอกาสที่จะหมุนแล้วมาตกที่ช่องมีมากขึ้น ในทางกลับกัน โครโมโซมที่มีค่าความเหมาะสมน้อยจะมีความน่าจะเป็นที่ถูกสุ่มเลือกได้สูง ดังนั้นช่องสำหรับโครโมโซมก็มีขนาดแคบซึ่งทำให้โอกาสที่จะหมุนแล้วมาตกที่ช่องมีน้อยลง โดยขั้นตอนการทำงานมายละเอียดดังนี้

1. คำนวณค่าความเหมาะสมของแต่ละโครโมโซมในประชากรเริ่มต้น ด้วยฟังก์ชั่นคำนวณค่าความเหมาะสม ดังแสดงตัวอย่างในรูปที่ 3.16 ในช่อง "$\mathbb{F}(S)$"

2. ทำการรวมค่าความเหมาะสมของแต่ละโครโมโซมแบบสะสม (Accumulative) จากตัวแรกไปถึงตัวสุดท้าย ดังแสดงตัวอย่างในรูปที่ 3.16 ในช่อง "Acc.

3. ทำการนอมอลไลซ์ โดยหารด้วยค่าความเหมาะสมสะสมรวม ซึ่งจะทำให้ได้ค่าความเหมาะสมที่อยู่ในช่วง 0 ถึง 1 ดังแสดงตัวอย่างในรูปที่ 3.16 ในช่อง "Norm"

4. พิสัยของค่าความเหมาะสมที่ทำการนอมอลไลซ์แล้วก็คือค่าความน่าจะเป็นของแต่ละโครโมโซม ดังแสดงตัวอย่างในรูปที่ 3.16 ในช่อง "Prob"

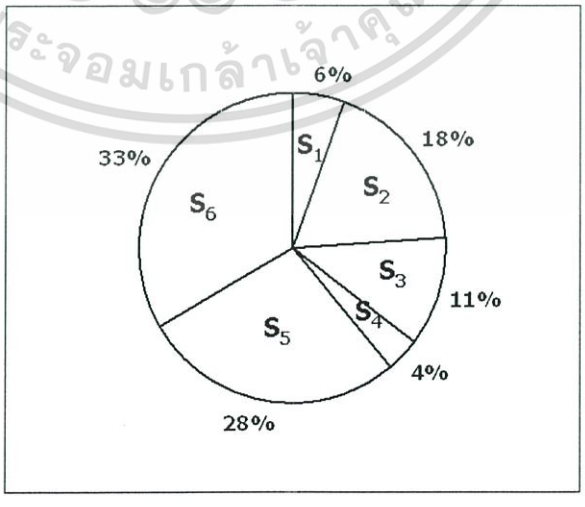
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Pop	$\mathbb{F}(S)$	Acc	Norm	Prob
S_1	0.15	0.15	0.056	0.056
S_2	0.50	0.65	0.240	0.184
S_3	0.30	0.95	0.352	0.112
S_4	0.10	1.05	0.389	0.037
S_5	0.75	1.80	0.667	0.278
S_6	0.90	2.70	1	0.333

รูปที่ 3.16 ตัวอย่างขั้นตอนการสร้างรูเล็ทวีล

จากรูปที่ 3.16 แสดงตัวอย่างขั้นตอนการสร้างรูเล็ทวีลของประชากรที่ประกอบไปด้วยโครโมโซม 6 ตัว คือ S_1 ถึง S_6 แสดงในช่อง "Pop" จากนั้นทำการคำนวณค่าความเหมาะสมของแต่ละโครโมโซม แสดงในช่อง $\mathbb{F}(S)$ ต่อมาทำการรวมค่าความเหมาะสมของแต่ละโครโมโซมแบบสะสมจากตัวแรกไปถึงตัวสุดท้าย แสดงในช่อง "Acc" จากนั้นทำการนอมอลไลซ์เพื่อให้ค่าความเหมาะสมอยู่ในช่วง 0 ถึง 1 แสดงในช่อง "Norm" และสุดท้ายจะได้พิสัยของค่าความเหมาะสมที่ทำการนอมอลไลซ์แล้ว ซึ่งก็คือค่าความน่าจะเป็นของแต่ละโครโมโซม เช่นโครโมโซม S_1 จะมีพิสัยกว้าง 0 ถึง 0.056 ค่าความน่าจะเป็นจะเท่ากับ 0.056 หรือโครโมโซม S_2 จะพิสัยมีกว้าง 0.056 ถึง 0.240 ค่าความน่าจะเป็นจะเท่ากับ 0.184 เป็นต้น ดังแสดงเป็นแผนภูมิวงกลมในรูปที่ 3.17

Prob
0.056
0.184
0.112
0.037
0.278
0.333



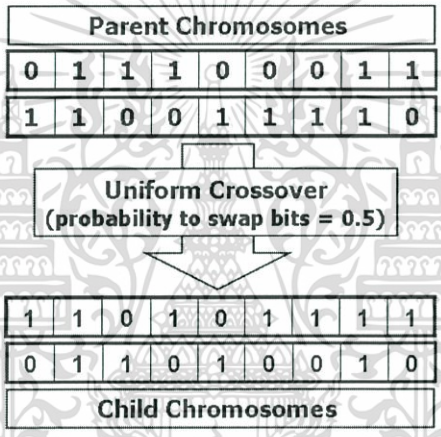
รูปที่ 3.17 แผนภูมิวงกลมของตัวอย่างการสร้างรูเล็ทวีล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **การสุ่มเลือกโครโมโซมพ่อ-แม่ (Parent Selection)**

การสุ่มเลือกโครโมโซมพ่อ-แม่เป็นการที่จะสุ่มเลือกโครโมโซมจากกลุ่มประชากรโดยใช้หลักรูเล็ทวีล ขึ้นมาจำนวน 2 ตัว (ที่ไม่ซ้ำกัน) เพื่อใช้เป็นโครโมโซมพ่อ-แม่ (Parent Chromosomes) จากนั้นจะนำโครโมโซมเหล่านั้นมาผสมกันด้วยกระบวนการครอสโอเวอร์ ซึ่งการครอสโอเวอร์ก็มีมากมายหลายวิธี โดยในวิทยานิพนธ์ฉบับนี้จะใช้ "Uniform Crossover" โดยจะทำการไล่ไปทีละบิตระหว่างโครโมโซมพ่อ-แม่ จากนั้นสุ่มสลับค่าบิตด้วยความน่าจะเป็น $P_c = 0.5$

การทำกระบวนการครอสโอเวอร์0 1 ครั้งจะทำให้ได้โครโมโซมลูก (Child Chromosomes) จำนวน 2 ตัว โดยในวิทยานิพนธ์ฉบับนี้กำหนดให้ในแต่ละเจเนอเรชั่นจะทำกระบวนการ Crossover จำนวน $N_{pop}/2$ ครั้ง ซึ่งจะทำให้ได้โครโมโซมลูกทั้งหมดจำนวน $N_{pop}/2$ ตัว ดังแสดงตัวอย่างในรูปที่ 3.18



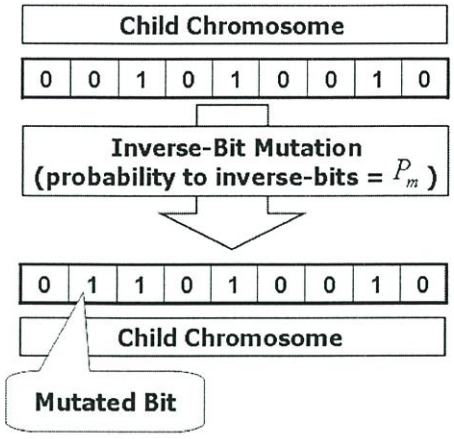
รูปที่ 3.18 ตัวอย่างการทำงานของกระบวนการครอสโอเวอร์ของ GAs

จากรูปที่ 3.18 แสดงตัวอย่างการทำงานของกระบวนการครอสโอเวอร์ของ GAs ซึ่งเป็นแบบ "Uniform Crossover" โดยที่โครโมโซมพ่อ-แม่ มีขนาด 9 บิต ซึ่งจะเห็นว่าบิตที่ 1, 3, 6, และ 7 ของโครโมโซมพ่อ-แม่จะถูกสลับกัน โดยการทำกระบวนการครอสโอเวอร์ 1 ครั้งจะทำให้เกิดโครโมโซมลูกออกมาจำนวน 2 ตัว

3.2.2.3 กระบวนการกลายพันธุ์ (Mutation Process)

กระบวนการกลายพันธุ์เป็นกระบวนการที่จะทำการเปลี่ยนแปลงค่าบิตในโครโมโซม ซึ่งการกลายพันธุ์มีมากมายหลายวิธีเช่นกัน โดยในวิทยานิพนธ์ฉบับนี้จะใช้ "Inverse-Bit Mutation" ซึ่งเป็นการไล่ไปทีละบิตในโครโมโซมลูกหลาน จากนั้นสุ่มเปลี่ยนค่าบิตในโครโมโซมลูกหลาน โดยจะเปลี่ยนค่าจาก "0" เป็น "1" หรือจาก "1" เป็น "0" (กลายพันธุ์) ด้วยค่าความน่าจะเป็น P_m ตามที่ได้กำหนดไว้ล่วงหน้า ดังแสดงตัวอย่างในรูปที่ 3.19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



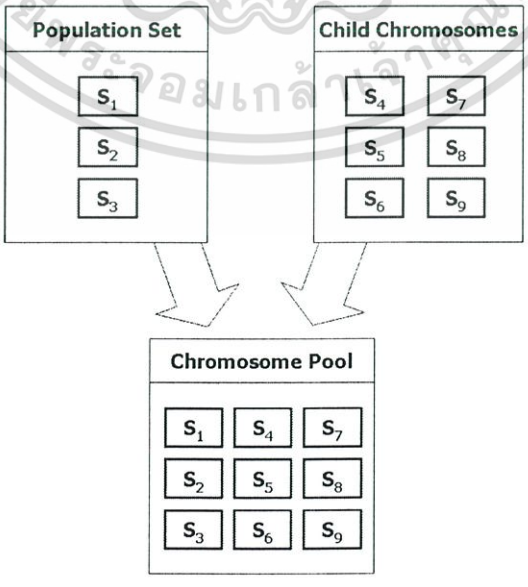
รูปที่ 3.19 ตัวอย่างการทำงานของกระบวนการกลายพันธุ์ของ GAs

3.2.2.4 กระบวนการเลือกสรรประชากรรุ่นถัดไป (Selection Process)

กระบวนการเลือกสรรประชากรรุ่นถัดไปเป็นกระบวนการที่จะเลือกเก็บโครโมโซมบางส่วนจากโครโมโซมที่มีอยู่ทั้งหมดไปเป็นประชากรโครโมโซมในรุ่นถัดไป และทำการปรับปรุงโครโมโซมที่ดีที่สุดด้วย โดยในการเลือกโครโมโซมจะเลือกโครโมโซมโซมที่มีค่าความเหมาะสมมากที่สุดก่อน และจะทำการเลือกเป็นจำนวน N_{pop} ตัว โดยมีขั้นตอนการทำงานมีรายละเอียดดังนี้

- **การนำโครโมโซมทั้งหมดมารวมกัน (Chromosome Pooling)**

การนำโครโมโซมทั้งหมดมารวมกันเป็นการที่นำโครโมโซมในประชากร กับโครโมโซมลูกที่ได้จากกระบวนการครอสโอเวอร์ และกระบวนการกลายพันธุ์แล้ว มารวมกันไว้เป็นกลุ่มโครโมโซมร่วม (Chromosome Pool) ดังแสดงตัวอย่างในรูปที่ 3.20



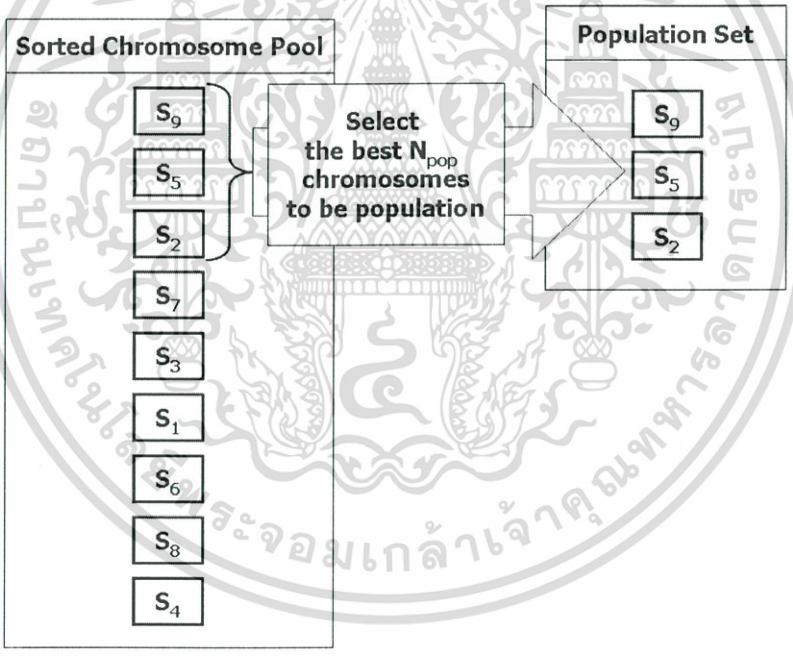
รูปที่ 3.20 ตัวอย่างการสร้างกลุ่มโครโมโซมร่วมของ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.20 แสดงตัวอย่างการสร้างโครโมโซมกองกลาง โดยกำหนดให้ประชากรมีสมาชิกเป็นโครโมโซมจำนวน 3 ตัว (S_1 ถึง S_3) และมีโครโมโซมลูกจำนวน 6 ตัว (S_4 ถึง S_9) เมื่อนำมารวมกันก็จะได้กลุ่มโครโมโซมร่วมจำนวน 9 ตัว

- **การเลือกโครโมโซม (Chromosome Selection)**

การเลือกโครโมโซมเป็นกระบวนการที่ทำการเลือกโครโมโซมที่อยู่ในกลุ่มโครโมโซมร่วม เพื่อนำไปเป็นประชากรในรุ่นถัดไป ซึ่งสามารถทำได้โดยการนำโครโมโซมในกลุ่มโครโมโซมร่วมมาคำนวณหาค่าความเหมาะสมโดยใช้ฟังก์ชั่นคำนวณค่าความเหมาะสม จากนั้นทำการเรียงลำดับโครโมโซมทั้งหมดในกลุ่มโครโมโซมร่วมตามค่าความเหมาะสม จากมากสุดไปน้อยสุด ซึ่งการทำเช่นนี้จะเรียกว่า "Elitist Strategy" จากนั้นจะทำการเลือกโครโมโซมที่มีค่าความเหมาะสมมากที่สุดจำนวน N_{pop} ตัว ไปเป็นประชากรในรุ่นถัดไป ดังแสดงตัวอย่างในรูปที่ 3.21



รูปที่ 3.21 ตัวอย่างการเลือกสรรประชากรรุ่นถัดไปของ GAs

จากรูปที่ 3.21 แสดงตัวอย่างการเลือกสรรประชากรรุ่นถัดไป โดยกำหนดให้โครโมโซมที่เรียงลำดับตามค่าความเหมาะสมจากมากสุดไปน้อยสุดแสดงในช่องทางด้านซ้าย ซึ่งจะเห็นว่าโครโมโซม S_9, S_5 และ S_2 เป็นโครโมโซม 3 ตัวแรกที่มีค่าความเหมาะสมมากที่สุด ดังนั้นจึงเลือกโครโมโซม S_9, S_5 และ S_2 ไปเป็นโครโมโซม (เนื่องจากจำนวนประชากรมีค่าเท่ากับ 3 จึงทำการเลือกโครโมโซม 3 ตัวไปเป็นประชากรในรุ่นถัดไป)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **การปรับปรุงค่าโครโมโซมที่ดีที่สุด (Best Chromosome Updating)**

การปรับปรุงค่าโครโมโซมที่ดีที่สุดเป็นการที่ทำการปรับปรุงค่าโครโมโซมที่ดีที่สุด ซึ่งเป็นการเก็บโครโมโซมที่มีความสอดคล้องกับชุดข้อมูลฝึกสอนและมีจำนวนโปรโตไทป์น้อยกว่าโครโมโซมปัจจุบันที่ดีที่สุด ซึ่งสามารถได้โดยการพิจารณากลุ่มโครโมโซมร่วมทีละโครโมโซมว่ามีโครโมโซมใดบ้างที่มีค่า $Accuracy(S)$ (จำนวนออบเจ็คในชุดข้อมูลฝึกสอนที่ถูกจำแนกโดยใช้ NN Rule ที่มีโปรโตไทป์ S เป็นข้อมูลอ้างอิงได้อย่างถูกต้อง) เท่ากับ $N_TrainingDataset$ (จำนวนออบเจ็คทั้งหมดของชุดข้อมูลฝึกสอน) ถ้าเท่ากันแสดงว่าโครโมโซมนั้นมีความสอดคล้องกับชุดข้อมูลฝึกสอนจากนั้นพิจารณาต่อว่าโครโมโซมนั้นๆ มีค่า $size(S)$ (จำนวนออบเจ็คที่ถูกเลือกเป็นโปรโตไทป์ของโปรโตไทป์ S) น้อยกว่าโครโมโซมปัจจุบันที่ดีที่สุดหรือไม่ ถ้าน้อยกว่าจะกำหนดให้โครโมโซมที่ดีที่สุดมีค่าเท่ากับโครโมโซมนั้น แต่ถ้ามากกว่าก็จะไปพิจารณาโครโมโซมตัวถัดไป

จากขั้นตอนการทำงานของ GAs ที่ได้อธิบายไปจะเห็นว่า GAs เป็นการค้นหาคำตอบแบบกว้างที่อาศัยกระบวนการทางพันธุกรรมในการเฟ้นสุ่มค้นหาคำตอบบนปริภูมิคำตอบ ดังนั้น GAs จึงสามารถนำมาประยุกต์ใช้ในการเลือกโปรโตไทป์ได้ดีเฉพาะกับชุดข้อมูลฝึกสอนที่มีข้อมูลจำนวนน้อยๆ เท่านั้น เพราะในกรณีของชุดข้อมูลฝึกสอนที่มีข้อมูลจำนวนน้อยๆ จะมีปริภูมิคำตอบขนาดไม่ใหญ่มาก GAs เพียงอย่างเดียวก็สามารถเฟ้นสุ่มหาคำตอบที่ดีเจอ แต่เมื่อชุดข้อมูลฝึกสอนมีข้อมูลจำนวนมากขึ้น เช่นชุดข้อมูลฝึกสอนที่มีข้อมูลจำนวน 1000 ตัว ก็จะมีปริภูมิคำตอบขนาด $2^{1000} = 1.071x10^{301}$ จึงเป็นไปได้ยากที่จะ GAs เพียงอย่างเดียวจะสามารถเฟ้นสุ่มหาคำตอบที่ดีเจอ ซึ่งเป็นข้อด้อยของ GAs

จึงเป็นที่มาของปัญหาในวิทยานิพนธ์ฉบับนี้ที่จะต้องหาความรู้ที่ดีมาช่วยไกด์ ที่ทำให้ GAs สามารถเฟ้นสุ่มหาคำตอบที่ดีได้แม้ในชุดข้อมูลที่ข้อมูลจำนวนมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม

จากปัญหาของวิธี MCS ที่แสดงให้เห็นในบทที่ 3 ว่าการเริ่มต้นการทำงานของวิธี MCS โดยใช้ออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนในการคำนวณความครอบคลุมให้ผลลัพธ์ที่ไม่ใช่ผลลัพธ์ที่ดีที่สุด (Global Optimum) แต่เป็นเพียงผลลัพธ์ที่ดีที่สุดเฉพาะที่ (Local Optimum) และตามที่แสดงให้เห็นในบทที่ 3 เช่นกันว่า ถ้าใช้แค่บางออบเจ็คในชุดข้อมูลฝึกสอนในการเริ่มต้นการทำงาน อาจจะทำให้ได้โปรโตไทป์ผลลัพธ์ที่ดีขึ้นกว่าเดิม

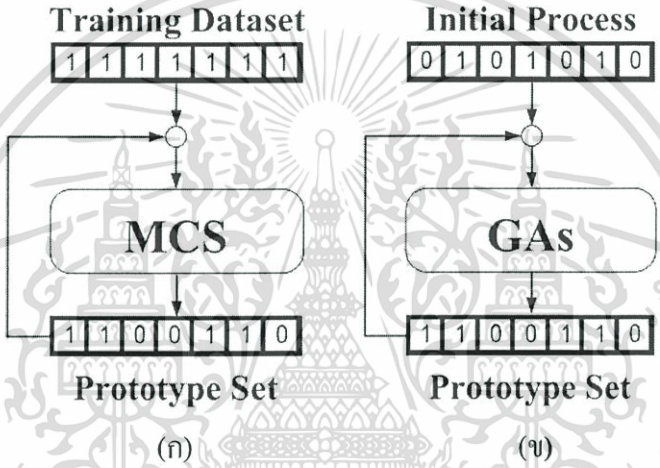
จึงเป็นปัญหาของวิทยานิพนธ์ฉบับนี้ที่จะต้องหาวิธีกำหนดออบเจ็คที่ใช้ในการเริ่มต้นการทำงานของวิธี MCS เพื่อที่จะให้ได้ผลลัพธ์ที่ดีขึ้นกว่าเดิม ซึ่งปัญหานี้เป็นปัญหาแบบ "ขัดแย้งกันเอง" กล่าวคือการที่จะทราบว่าควรกำหนดออบเจ็คใดบ้างในการเริ่มต้นการทำงาน ก็ต้องทราบก่อนว่ามีออบเจ็คใดบ้างที่น่าจะเลือกเป็นโปรโตไทป์บ้าง ซึ่งก็คือผลลัพธ์ของการเลือกโปรโตไทป์นั่นเอง อีกทั้งข้อมูลในแต่ละชุดข้อมูลก็มีลักษณะการวางตัวบนปริภูมิคุณลักษณะเฉพาะที่แตกต่างกันออกไป ทำให้ไม่มีวิธีที่สามารถบอกได้แน่นอนว่าควรจะพิจารณาใช้ หรือไม่ใช้ออบเจ็คใดบ้างในการเริ่มต้นการทำงานของวิธี MCS ดังนั้นวิทยานิพนธ์ฉบับนี้จึงเสนอการนำ GAs มาช่วยในการกำหนดออบเจ็คที่ใช้ในการเริ่มต้นการทำงานของวิธี MCS

ส่วนในปัญหาของ GAs ตามที่กล่าวไว้ในบทที่ 3 เช่นกัน ว่าสามารถนำมาประยุกต์ใช้ในการเลือกโปรโตไทป์ได้ดีเฉพาะกับชุดข้อมูลฝึกสอนที่มีจำนวนข้อมูลน้อยๆ เท่านั้น เพราะ GAs เป็นการค้นหาคำตอบที่อาศัยกระบวนการทางพันธุกรรมในการค้นหาคำตอบ (ซึ่งกระบวนการทางพันธุกรรมเป็นวิธีการเชิงสุ่ม) โดยในกรณีชุดข้อมูลฝึกสอนที่มีข้อมูลจำนวนน้อยๆ ก็จะมีปริภูมิคำตอบขนาดไม่ใหญ่มาก ทำให้ GAs เพียงอย่างเดียวสามารถเฟ้นสุ่มหาคำตอบที่ดีได้ แต่เมื่อชุดข้อมูลฝึกสอนมีข้อมูลจำนวนมากขึ้น เช่นชุดข้อมูลฝึกสอนที่มีข้อมูลจำนวน 1000 ตัว ก็จะมีปริภูมิคำตอบขนาด $2^{1000} = 1.071x10^{301}$ จึงเป็นไปได้ยากที่ GAs เพียงอย่างเดียวจะสามารถเฟ้นสุ่มหาคำตอบที่ดีได้ วิทยานิพนธ์ฉบับนี้จึงเสนอการนำวิธี MCS เข้ามาช่วยไกด์ ทำให้ GAs สามารถเฟ้นสุ่มหาคำตอบที่ดีได้ แม้ชุดข้อมูลฝึกสอนจะมีข้อมูลจำนวนมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1 ภาพรวมการทำงานของระบบ

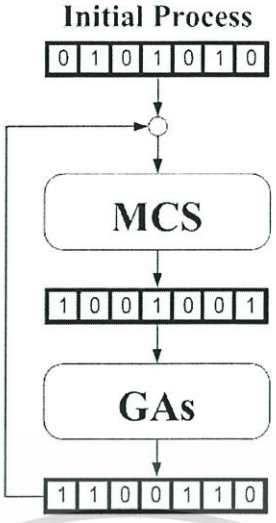
วิทยานิพนธ์ฉบับนี้ขอนำเสนอ “Hybrid Genetic Algorithms (HGAs)” ซึ่งเป็นการเลือกโปรโตไทป์โดยใช้วิธี MCS มาทำงานร่วมกับ GAs โดยเป็นการนำข้อดีของแต่ละวิธีมาแก้ข้อด้อยของอีกวิธี กล่าวคือจากข้อด้อยของวิธี MCS ที่มีการเริ่มต้นการทำงานที่ตายตัวที่อาจจะส่งผลให้ผลลัพธ์ที่ได้ไม่ใช่ผลลัพธ์ที่ดีที่สุด จึงนำ GAs เข้ามาช่วยสุ่มกำหนดออบเจ็คที่ใช้ในการเริ่มต้นการทำงานแทน ในทางกลับกันจากข้อด้อยของ GAs ที่ให้ผลลัพธ์ไม่ดีในข้อมูลฝึกสอนที่มีจำนวนข้อมูลมาก เนื่องมาจากปริภูมิคำตอบมีขนาดมหาศาล จึงเป็นไปได้ยากที่จะเฟ้นสุ่มหาคำตอบที่ดีได้ จึงนำความรู้จากหลัก “ความครอบคลุม” ของวิธี MCS มาช่วยไกด์ให้กับ GAs



รูปที่ 4.1 แผนภาพการทำงานของวิธี MCS และ GAs

รูปที่ 4.1 แสดงแผนภาพการทำงานของวิธี MCS และ GAs โดยรูป (ก) แสดงแผนภาพการทำงานของวิธี MCS ที่รับอินพุตเป็นข้อมูลฝึกสอน โดยเริ่มต้นการทำงานโดยการนำออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนมาคำนวณความครอบคลุม (แสดงในรูปของโครโมโซมที่กำหนดให้ทุกบิตมีค่าเป็น “1”) จากนั้นนำความรู้ที่ได้มาทำกรองโปรโตไทป์ที่ซ้ำซ้อนออก จนได้เอาพุตเป็นโปรโตไทป์ผลลัพธ์ จากนั้นนำโปรโตไทป์ผลลัพธ์ที่ได้ไปเป็นอินพุตในการทำงานรอบต่อไป และวนทำการกรองจนกว่าจำนวนโปรโตไทป์ผลลัพธ์จะมีจำนวนไม่ลดลง ส่วนรูป (ข) แสดงแผนภาพการทำงานของ GAs ที่มีกระบวนการกำหนดค่าเริ่มต้นในการสุ่มประชากรเริ่มต้นเป็นอินพุต จากนั้นนำประชากรเหล่านี้มาผ่านกระบวนการทางพันธุกรรมจนได้โปรโตไทป์ผลลัพธ์เป็นเอาพุต จากนั้นนำโปรโตไทป์ผลลัพธ์ที่ได้ไปเป็นประชากรเริ่มต้นในรอบถัดไป และวนทำงานเป็นรอบๆ จนกว่าจะตรงตามเงื่อนไขการสิ้นสุดการทำงานที่กำหนดไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 แผนภาพการทำงานของ HGAs

จากรูปที่ 4.2 แสดงแผนภาพการทำงานของ HGAs ที่นำวิธี MCS มาทำงานร่วมกับ GAs โดยเริ่มต้นการทำงานด้วยกระบวนการกำหนดค่าเริ่มต้นของ GAs ซึ่งเป็นการสุ่มกำหนดค่าให้กับประชากรเริ่มต้น จากนั้นนำประชากรเริ่มต้นมาเป็นอินพุตให้กับกระบวนการ MCS (เปรียบเสมือนการกำหนดออบเจ็คที่ใช้ในการคำนวณความครอบคลุมให้กับวิธี MCS) ต่อมานำโปรโตไทป์ผลลัพธ์ที่ได้จากกระบวนการ MCS ไปเป็นประชากรเริ่มต้นของ GAs จนได้โปรโตไทป์ผลลัพธ์เป็นเอาพุต แต่แทนที่จะนำโปรโตไทป์ผลลัพธ์ไปเป็นประชากรเริ่มต้นในรอบถัดไปโดยตรง ก็จะนำโปรโตไทป์ผลลัพธ์เหล่านี้ไปผ่านกระบวนการ MCS ก่อนที่จะนำไปเป็นประชากรเริ่มต้นในรอบถัดไป (เปรียบเสมือนการไกด์คำตอบให้กับ GAs) และวนทำงานเป็นรอบๆ จนกว่าจะตรงตามเงื่อนไขการสิ้นสุดการทำงานที่กำหนดไว้

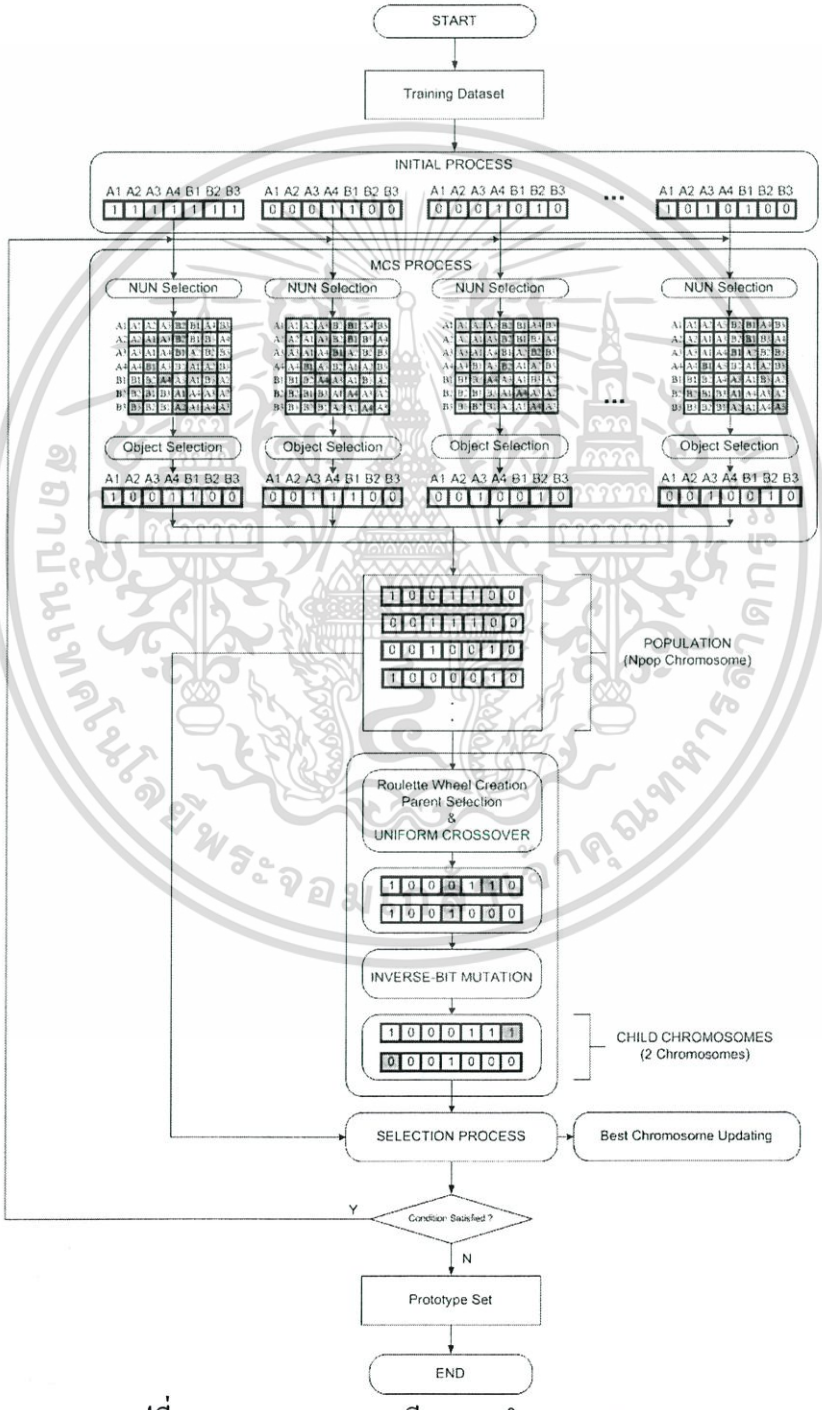
จากที่ได้กล่าวไปข้างต้นว่า HGAs เป็นการทำงานร่วมกันระหว่างวิธี MCS และ GAs โดยจะทำวิธี MCS ก่อนแล้วจึงทำ GAs หรือจะทำ GAs ก่อนแล้วจึงทำวิธี MCS ก็ได้ (รายละเอียดการทดลองแสดงในภาคผนวก ซ.) แต่เมื่อพิจารณาจากผลการทดลองในภาคผนวก ซ. ซึ่งสรุปได้ว่า HGAs แบบที่ทำวิธี MCS ก่อนแล้วจึงทำ GAs ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ทำ GAs ดังนั้นวิทยานิพนธ์ฉบับนี้จึงนำเสนอ HGAs แบบที่ทำวิธี MCS ก่อนแล้วจึงทำ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยการทำงานของ HGAs สามารถพิจารณาได้ 2 มุมมองดังนี้

มุมมองแรก คือ HGAs เปรียบเสมือนวิธี MCS ที่มี GAs ช่วยในการสุ่มออบเจ็คที่ใช้ในการคำนวณคามครอบคลุม ทำให้ได้ผลลัพธ์ที่มีความหลากหลาย ซึ่งมีโอกาสที่ได้ผลลัพธ์ที่ดีกว่าเดิม

มุมมองที่สอง คือ HGAs เปรียบเสมือน GAs ที่มีวิธี MCS เป็นกระบวนการตัดต่อพันธุกรรม ทำให้เกิดการวิวัฒนาการแบบก้าวกระโดด ส่งผลให้ประชากรพัฒนาไปในทิศทางที่ถูกต้อง



รูปที่ 4.3 แผนภาพรายละเอียดการทำงานของ HGAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 รายละเอียดขั้นตอนการทำงาน

ขั้นตอนการทำงานของ HGAs พื้นฐานมาจาก GAs แบบดั้งเดิมตามที่ได้อธิบายไปในหัวข้อที่ 3.2 แต่จะมีการนำวิธี MCS เข้ามาทำงานร่วมด้วย อีกทั้งยังมีการนำเทคนิค "Singleton" เข้ามาช่วยเพิ่มประสิทธิภาพการทำงานในส่วนของ MCS และนำหลักการ "Diversity" เข้ามาช่วยเพิ่มประสิทธิภาพการทำงานในส่วนของ GAs อีกด้วย (ซึ่งจะอธิบายในหัวข้อต่อไป)

ตามที่ได้กล่าวไปข้างต้นว่า HGAs นั้นมีพื้นฐานมาจาก GAs แบบดั้งเดิม ดังนั้นกระบวนการพื้นฐานต่างๆ ก็จะเหมือนกับ GAs ดั้งเดิม กล่าวคือมีการเข้ารหัสโครโมโซม และมีฟังก์ชั่นคำนวณค่าความเหมาะสมเหมือนกันตามที่ได้อธิบายไปในหัวข้อที่ 3.2.1.1 และ 3.2.1.8 ตามลำดับ

รายละเอียดขั้นตอนการทำงานของ HGAs มีดังนี้

4.2.1 กระบวนการสร้างตารางระยะทาง (Distance Calculating Process)

กระบวนการสร้างตารางระยะทางของ HGAs มีพื้นฐานเดียวกันกับกระบวนการสร้างตารางระยะทางตามที่ได้อธิบายไปในหัวข้อที่ 3.1.2.1 กล่าวคือเป็นกระบวนการคำนวณระยะห่างระหว่าง 2 ออบเจ็คที่เป็นไปได้ทั้งหมดในชุดข้อมูลฝึกสอน โดยใช้ฟังก์ชั่นคำนวณระยะทาง จากนั้นนำค่าระยะห่างที่ได้มาเก็บไว้ในตารางระยะทาง เพื่อนำไปใช้ในการคำนวณความครอบคุลมต่อไป

4.2.2 กระบวนการกำหนดค่าเริ่มต้น (Initial Process)

กระบวนการกำหนดค่าเริ่มต้นของ HGAs จะมีพื้นฐานเดียวกันกับกระบวนการกำหนดค่าเริ่มต้นตามที่ได้อธิบายไปในหัวข้อที่ 3.2.2.1 กล่าวคือเป็นกระบวนการที่จะสร้างโครโมโซมตามจำนวนจำนวนประชากรตามที่กำหนดไว้ล่วงหน้า และทำการสุ่มกำหนดค่าให้กับโครโมโซมด้วยค่าความน่าจะเป็น (P_{ini}) ตามที่ได้กำหนดไว้ก่อนเช่นกัน โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

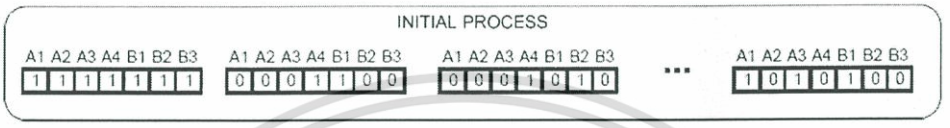
4.2.2.1 การสร้างประชากรเริ่มต้น (Population Creation)

การสร้างประชากรเริ่มต้นเป็นการสร้างโครโมโซมตามจำนวนเท่ากับจำนวนประชากรตามที่กำหนดไว้ล่วงหน้า โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.1-การสร้างประชากรเริ่มต้น"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2.2 การสุ่มกำหนดค่าให้กับประชากรเริ่มต้น (Population Initialization)

การสุ่มกำหนดค่าให้กับประชากรเริ่มต้นเป็นการสุ่มกำหนดค่าให้กับแต่ละบิตของโครโมโซมในประชากร ด้วยความน่าจะเป็น P_{ini} ตามที่ได้กำหนดไว้ล่วงหน้า โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.1-การสุ่มกำหนดค่าให้กับประชากร" แต่จะแตกต่างจากเดิมตรงที่จะทำการสุ่มกำหนดค่าให้กับแต่ละบิตของโครโมโซมให้กับโครโมโซมจำนวน $N_{pop}-1$ ตัวเท่านั้น ส่วนอีก 1 โครโมโซมจะทำการกำหนดค่าให้เป็น "1" ทั้งหมด ซึ่งก็คือการกำหนดออบเจ็คแบบวิธี MCS ดั้งเดิมนั่นเอง



รูปที่ 4.4 ตัวอย่างประชากรเริ่มต้นที่ได้จากกระบวนการกำหนดค่าเริ่มต้นของ HGAs

จากรูปที่ 4.4 แสดงตัวอย่างประชากรเริ่มต้นที่ได้จากกระบวนการกำหนดค่าเริ่มต้น ซึ่งจะเห็นว่าโครโมโซมแรกจะถูกกำหนดให้แต่ละบิตมีค่าเป็น "1" ส่วนโครโมโซมที่เหลือก็จะทำการสุ่มกำหนดค่าตามความน่าจะเป็น P_{ini} ตามที่ได้กำหนดไว้ล่วงหน้า

4.2.2.3 การสร้างโครโมโซมที่ดีที่สุด (The Best Chromosome Creation)

การสร้างโครโมโซมที่ดีที่สุดเป็นการสร้างโครโมโซมที่ดีที่สุด โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.1-การสร้างโครโมโซมที่ดีที่สุด" ซึ่งในวิทยานิพนธ์ฉบับนี้จะนิยามโครโมโซมที่ดีที่สุดต้องเป็นไปตามเงื่อนไข 2 ข้อ ดังนี้

เงื่อนไขที่ 1 คือ โครโมโซมจะต้องมีความสอดคล้องกับชุดข้อมูลฝึกสอน

เงื่อนไขที่ 2 คือ โครโมโซมจะต้องมีค่า *Size*(S) น้อยที่สุดในบรรดาโครโมโซมทั้งหมดที่ได้เคยมีการสร้างขึ้นมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 กระบวนการ MCS (MCS Process)

กระบวนการ MCS ของ HGAs มีพื้นฐานเดียวกันกับวิธี MCS แบบเดิมตามที่ได้อธิบายไปในหัวข้อที่ 3.1 กล่าวคือจะทำการคำนวณค่าความสามารถในการครอบคลุม และทำการเลือกโปรโตไทป์โดยพิจารณาจากค่าความสามารถในการครอบคลุมนั้น แต่จะต่างจากเดิมตรงที่กระบวนการ MCS จะงานเพียงรอบเดียวไม่เหมือนกับวิธี MCS ที่จะวนรอบการทำงานจนกว่าจำนวนโปรโตไทป์ผลลัพธ์จะมีจำนวนไม่ลดลง โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

4.2.3.1 กระบวนการคำนวณความครอบคลุม (Cover Calculating Process)

กระบวนการสร้างตารางความครอบคลุมเป็นกระบวนการที่จะคำนวณความครอบคลุม เพื่อใช้ในการเลือกโปรโตไทป์ในขั้นตอนต่อไป โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ 3.1.2.2 ดังนี้

- **การเรียงลำดับระยะทาง (Distance Sorting)**

การเรียงลำดับระยะทางเป็นการนำข้อมูลในตารางระยะทางของแต่ละออบเจ็คมาเรียงลำดับจากน้อยไปหามาก โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.1.2.2-การเรียงลำดับระยะทาง"

- **การเลือกออบเจ็คต่างชนิด (NUN Selection)**

การเลือกออบเจ็คต่างชนิดเป็นการนำข้อมูลในตารางระยะทางที่เรียงลำดับแล้วมาพิจารณาเลือก NUN โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.1.2.2-การเลือกออบเจ็คต่างชนิด"

- **การพิจารณาความครอบคลุม (Cover Determination)**

การพิจารณาความครอบคลุมเป็นการพิจารณาออบเจ็คที่สามารถครอบคลุมแต่ละออบเจ็คได้ โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.1.2.2-การพิจารณาความครอบคลุม"

- **การคำนวณค่าความสามารถในการครอบคลุม (Cover Value Calculation)**

การคำนวณค่าความสามารถในการครอบคลุมเป็นพิจารณาว่าแต่ละออบเจ็คสามารถครอบคลุมออบเจ็คในชุดข้อมูลฝึกสอนได้มากน้อยเพียงใด โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.1.2.2-การคำนวณค่าความสามารถในการครอบคลุม"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3.2 กระบวนการการเลือกออบเจ็ค (Object Selecting Process)

กระบวนการการเลือกออบเจ็คเป็นกระบวนการที่จะทำการเลือกออบเจ็คเป็นโปรโตไทป์ โดยพิจารณาจากค่าความสามารถในการครอบคลุมของแต่ละออบเจ็ค โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ 3.1.2.3 แต่จะแตกต่างจากเดิมตรงที่อาจจะต้องหยุดการเลือกแม้ว่าจะยังเหลือออบเจ็คที่ยังไม่ถูกครอบคลุมอยู่ก็ตาม (จากเดิมที่ทำการเลือกจนกว่าทุกออบเจ็คในชุดข้อมูลฝึกสอนถูกครอบคลุม) เนื่องมาจาก Candidate Set ที่ใช้ในกระบวนการ MCS ของ HGAs ไม่ได้พิจารณาจากทุกออบเจ็คในชุดข้อมูลฝึกสอนแต่มาจากการสุ่ม ดังนั้นจึงเป็นไปได้ยากที่โปรโตไทป์ผลลัพธ์ที่ได้จะครอบคลุมทุกออบเจ็คในชุดข้อมูลฝึกสอน ตัวอย่างเช่น จากตัวอย่างชุดข้อมูลที่ 2 (หัวข้อที่ 3.1.2 ในบทที่ 3) ถ้ากำหนด Candidate Set ตัวอย่างที่ 2 เป็น {A3, B1} สามารถทำการเลือกโปรโตไทป์ได้ผลลัพธ์ดังนี้

จากตัวอย่างตัวอย่างชุดข้อมูลที่ 2 เมื่อกำหนด Candidate Set ตาม Candidate Set ตัวอย่างที่ 2 จะได้ตารางค่าความสามารถในการครอบคลุมดังนี้

ตารางที่ 4.1 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2

Candidate Set = {A3, B1}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	0	0	0	0	0
A3	1	0	1	0	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	1	1	1
B2	0	0	0	0	1	1	1
B3	0	0	0	0	0	1	1
CV	3	2	2	1	2	3	3

การเลือกโปรโตไทป์ยังคงใช้หลักการเดิม กล่าวคือทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจาก Candidate Set เป็นโปรโตไทป์ ซึ่งก็คือออบเจ็ค A3 หรือออบเจ็ค B1 เพราะทั้ง 2 ออบเจ็คมีค่าความครอบคลุมสูงสุดเท่ากับ “2” และอยู่ใน Candidate Set โดยในที่นี้สมมติว่าเลือกออบเจ็ค B1 เป็นโปรโตไทป์ จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค B1 และ B2 ที่ถูกครอบคลุมโดยออบเจ็ค B1 แล้วออก ดังแสดงผลลัพธ์ในตารางที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B1 เป็นโปรโตไทป์แล้ว

Candidate Set = {A3, B1}
Prototype Set = {B1}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	0	0	0	0	0
A3	1	0	1	0	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	X	X	X
B2	0	0	0	0	X	X	X
B3	0	0	0	0	0	1	1
CV	3	2	2	1	0	1	1

ต่อมาทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจาก Candidate Set ลำดับต่อไปเป็นโปรโตไทป์ ซึ่งจากตารางที่ 4.2 จะเห็นว่าออบเจ็ค A3 มีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ "2" และอยู่ใน Candidate Set ดังนั้นจึงทำการเลือกออบเจ็ค A3 เป็นโปรโตไทป์ จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค A1 และ A3 ที่ถูกครอบคลุมโดยออบเจ็ค A3 แล้วออก ดังแสดงในตารางที่ 4.3

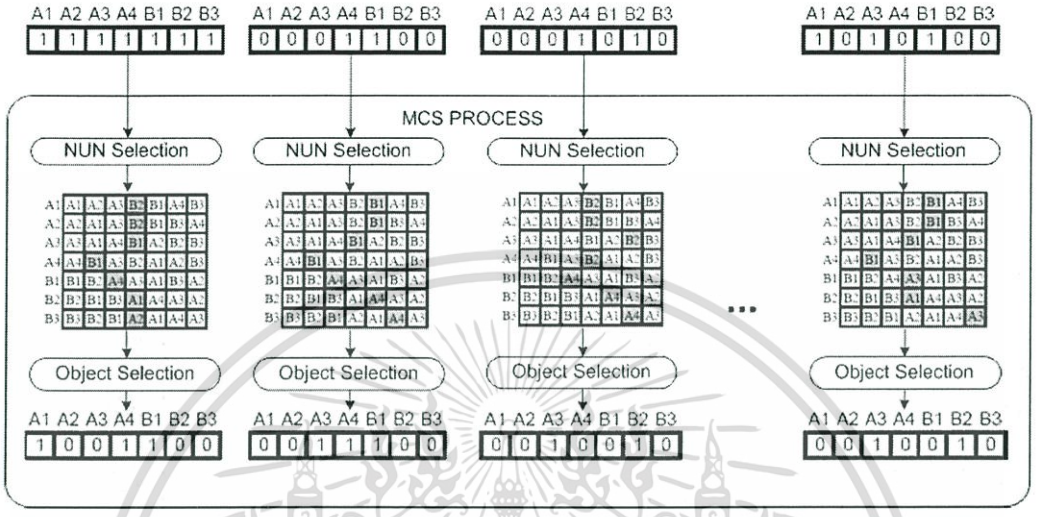
ตารางที่ 4.3 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค A3 และ B1 เป็นโปรโตไทป์แล้ว

Candidate Set = {A3, B1}
Prototype Set = {A3, B1}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	1	1	0	0	0	0	0
A3	X	0	X	0	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	X	X	X
B2	0	0	0	0	X	X	X
B3	0	0	0	0	0	1	1
CV	1	1	0	1	0	1	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

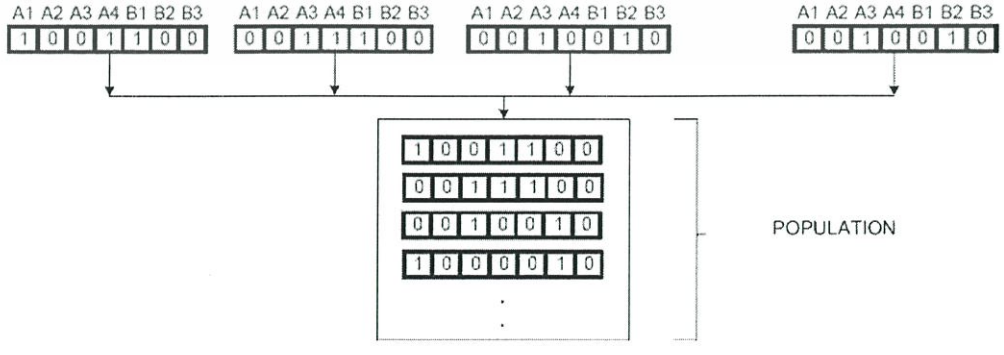
จากตารางที่ 4.3 จะเห็นว่าไม่มีออบเจ็คใดใน Candidate Set ให้เลือกแล้ว แต่ออบเจ็คในชุดข้อมูลที่ 2 ยังถูกครอบคลุมไม่หมด และไม่มีทางที่จะทำให้ทุกออบเจ็คในชุดข้อมูลฝึกสอนครอบคุลมได้ ดังนั้นจึงเปลี่ยนมาเป็นทำการเลือกจนกว่าจะไม่สามารถเลือกต่อได้ และต้องหยุดเลือกแม้ว่าจะยังมีออบเจ็คที่ยังไม่ถูกครอบคลุมอยู่ก็ตาม



รูปที่ 4.5 ตัวอย่างการทำงานของกระบวนการ MCS

จากรูปที่ 4.5 แสดงตัวอย่างการทำงานของกระบวนการ MCS ที่นำประชากรเริ่มต้นจากกระบวนการกำหนดค่าเริ่มต้นมาเป็น Candidate Set ในการเลือกโปรโตไทป์ โดยภายในกระบวนการ MCS ก็จะมีการคำนวณค่าความสามารถในการความครอบคลุม ซึ่งจะเห็นว่าการใช้ออบเจ็คในการคำนวณความครอบคลุมต่างกัน การเลือกออบเจ็คต่างชนิดก็จะให้ผลที่แตกต่างกัน ส่งผลให้โปรโตไทป์ผลลัพธ์ที่ได้แตกต่างกันตามไปด้วย

เมื่อผ่านกระบวนการ MCS แล้วจะได้ผลลัพธ์จำนวน N_{pop} ตัว ซึ่งผลลัพธ์เหล่านี้จะถูกนำไปเป็นประชากรเริ่มต้นของกระบวนการทางพันธุกรรมต่อไป ดังแสดงตัวอย่างในรูปที่ 4.6



รูปที่ 4.6 ตัวอย่างผลลัพธ์ที่ได้จากกระบวนการ MCS ของ HGAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4 กระบวนการครอสโอเวอร์ (Crossover Process)

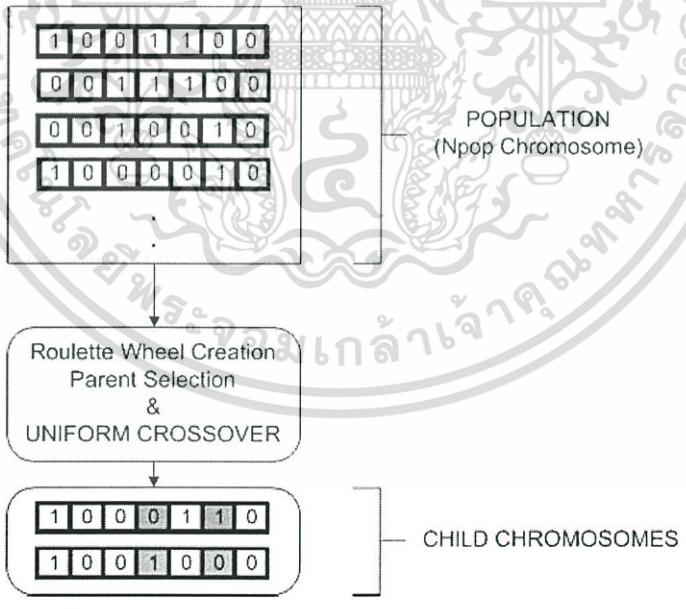
กระบวนการครอสโอเวอร์ของ HGAs มีพื้นฐานเดียวกันกับวิธี GAs ตามที่ได้อธิบายไปในหัวข้อที่ 3.2.2.2 กล่าวคือเป็นกระบวนการที่ทำการสุ่มเลือกโครโมโซมจากประชากรมาใช้เป็นพ่อ-แม่พันธุ์ เพื่อสร้างโครโมโซมลูกหลาน โดยใช้ "Uniform Crossover" โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

4.2.4.1 การสร้างรูเล็ทวีล (Roulette Wheel Creation)

การสร้างรูเล็ทวีล เป็นมีหลักการทำงานจาก "Roulette Wheel" ใช้เพื่อช่วยในการสุ่มเลือกโครโมโซมพ่อ-แม่ จากประชากรเริ่มต้นโดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.2-การสร้างรูเล็ทวีล"

4.2.4.2 การสุ่มเลือกโครโมโซมพ่อ-แม่ (Parent Selection)

การสุ่มเลือกโครโมโซมพ่อ-แม่เป็นการสุ่มเลือกโครโมโซมจากประชากรโดยใช้หลักรูเล็ทวีล ขึ้นมาจำนวนตัว 2 (ที่ไม่ซ้ำกัน) เพื่อใช้เป็นโครโมโซมพ่อ-แม่ โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.2-การสุ่มเลือกโครโมโซมพ่อ-แม่" ดังแสดงตัวอย่างในรูปที่ 4.7

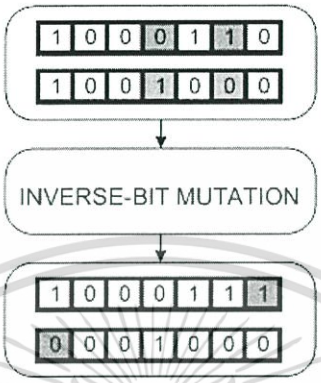


รูปที่ 4.7 ตัวอย่างการทำงานของกระบวนการครอสโอเวอร์ของ HGAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.5 กระบวนการกลายพันธุ์ (Mutation Process)

กระบวนการกลายพันธุ์ของ HGAs มีพื้นฐานเดียวกันกับวิธี GAs ตามที่ได้อธิบายไปในหัวข้อที่ 3.2.2.3 กล่าวคือเป็นกระบวนการที่จะทำการเปลี่ยนแปลงค่าบิตในโครโมโซม โดยใช้ "Inverse-Bit Mutation" ดังแสดงตัวอย่างในรูปที่ 4.8



รูปที่ 4.8 ตัวอย่างการทำงานของกระบวนการกลายพันธุ์ของ HGAs

4.2.6 กระบวนการเลือกสรรประชากรรุ่นถัดไป (Selection Process)

กระบวนการเลือกสรรประชากรรุ่นถัดไปของ HGAs มีพื้นฐานเดียวกันกับวิธี GAs ตามที่ได้อธิบายไปในหัวข้อที่ 3.2.2.4 กล่าวคือเป็นกระบวนการเลือกเก็บโครโมโซมบางส่วนจากโครโมโซมที่มีอยู่ทั้งหมดไปเป็นประชากรโครโมโซมในรุ่นถัดไป และทำการเก็บโครโมโซมที่ดีที่สุด โดยขั้นตอนการทำงานมีรายละเอียดดังนี้

4.2.6.1 การนำโครโมโซมทั้งหมดมารวมกัน (Chromosome Pooling)

การนำโครโมโซมทั้งหมดมารวมกันเป็นการนำโครโมโซมในประชากร กับโครโมโซมลูกที่ได้จากกระบวนการครอสโอเวอร์ และกระบวนการกลายพันธุ์แล้ว มารวมกันไว้เป็นกลุ่มโครโมโซมร่วม โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.4-การนำโครโมโซมทั้งหมดมารวมกัน"

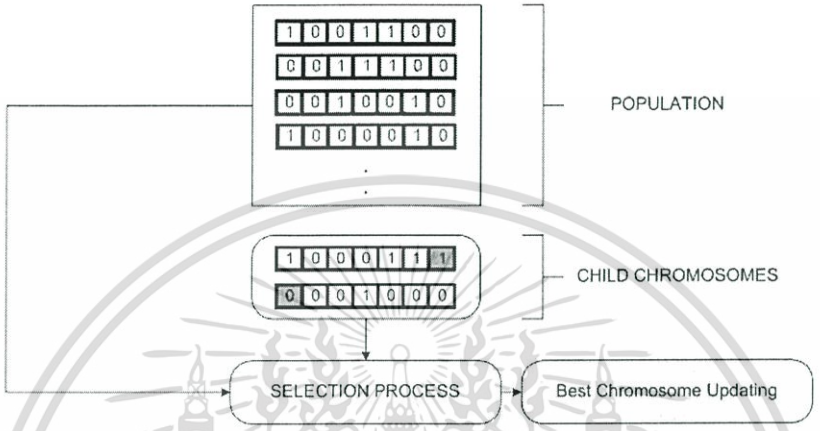
4.2.6.2 การเลือกโครโมโซมรุ่นถัดไป (Chromosome Selection)

การเลือกโครโมโซมรุ่นถัดไปเป็นการเลือกโครโมโซมที่อยู่ในกลุ่มโครโมโซมร่วม เพื่อนำไปเป็นประชากรในรุ่นถัดไป โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.4-การเลือกโครโมโซม"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.6.3 การปรับปรุงค่าโครโมโซมที่ดีที่สุด (The Best Chromosome Updating)

การปรับปรุงค่าโครโมโซมที่ดีที่สุดเป็นการปรับปรุงค่าโครโมโซมที่ดีที่สุด ซึ่งเป็นการเก็บโครโมโซมที่มีความสอดคล้องกับชุดข้อมูลฝึกสอน และมีจำนวนโปรโตไทป์น้อยกว่าโครโมโซมที่ดีที่สุดปัจจุบัน โดยมีรายละเอียดการทำงานเช่นเดียวกันกับหัวข้อที่ "3.2.2.4-การปรับปรุงค่าโครโมโซมที่ดีที่สุด"



รูปที่ 4.9 ตัวอย่างการทำงานของกระบวนการเลือกสรรประชากรในรุ่นถัดไปของ HGAs

จากรูปที่ 4.9 แสดงตัวอย่างการทำงานของกระบวนการเลือกสรรประชากรในรุ่นถัดไปของ HGAs แต่โครโมโซมที่เลือกได้จะไม่ได้นำไปเป็นประชากรเริ่มต้นในรอบถัดไปโดยตรง แต่จะนำโครโมโซมเหล่านี้ไปผ่านกระบวนการ MCS ก่อนที่จะนำไปเป็นประชากรเริ่มต้นในรอบถัดไป และจะวนทำงานจนกว่าจะครบตามเงื่อนไขที่กำหนดไว้ โดยในวิทยานิพนธ์ฉบับนี้กำหนดให้ HGAs ทำงานไปจนครบรอบตามเวลาที่กำหนดไว้ล่วงหน้า

4.3 การเพิ่มประสิทธิภาพการทำงานด้วยเทคนิค "Singleton"

จากแนวคิด "ความครอบคลุม" ของวิธี MCS ทำให้ทราบได้ว่าแต่ละออบเจ็คสามารถนำไปแทนออบเจ็คในชุดข้อมูลฝึกสอนได้มากน้อยเพียงใด โดยแสดงออกมาในรูปของค่าความสามารถในการครอบคลุม ทำให้สามารถเลือกออบเจ็คที่สามารถแทนออบเจ็คอื่นๆ ได้มากเป็นโปรโตไทป์ได้ แต่ถ้าลองพิจารณาที่ตารางความครอบคลุมซึ่งเป็นตารางที่แสดงออบเจ็คที่สามารถครอบคลุมแต่ละออบเจ็คได้ในอีกมุมมองหนึ่ง ก็จะทำให้ได้ความรู้อีกแบบ ตัวอย่างเช่น เมื่อพิจารณาที่ออบเจ็ค A4 จากตางรางที่ 3.4 ดังแสดงในรูปที่ 4.10

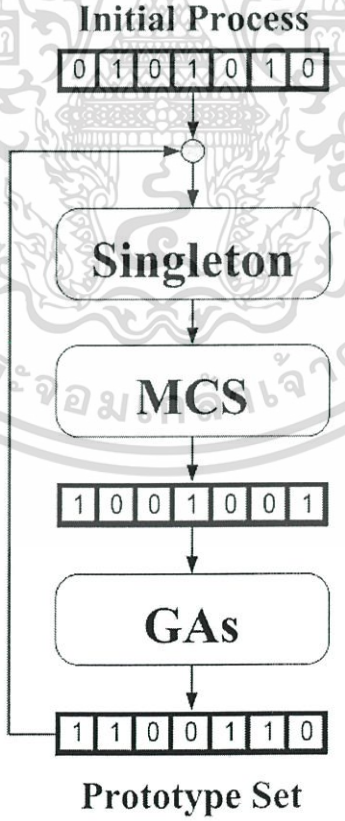
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	A1	A2	A3	A4	B1	B2	B3
A4	0	0	0	1	0	0	0

รูปที่ 4.10 ส่วนหนึ่งของตารางความรอบคลุมของออบเจ็ค A4 จากชุดข้อมูลตัวอย่างที่ 2

จากรูปที่ 4.9 แสดงส่วนหนึ่งของตารางความรอบคลุมของออบเจ็ค A4 จากชุดข้อมูลตัวอย่างที่ 2 ซึ่งจะเห็นว่าออบเจ็คที่สามารถครอบคลุมออบเจ็ค A4 ได้มีเพียงออบเจ็คเดียว ซึ่งก็คือออบเจ็ค A4 เอง นั่นหมายความว่าถ้าต้องการให้ออบเจ็ค A4 ถูกครอบคลุม ก็จะต้องเลือกออบเจ็ค A4 เองเป็นโปรโตไทป์เท่านั้น ซึ่งเรียกออบเจ็คที่มีลักษณะแบบนี้ว่า "Singleton"

ดังนั้นหมายความว่าถ้าต้องการให้ออบเจ็คที่มีลักษณะ Singleton ถูกครอบคลุมก็ต้องเลือกออบเจ็คนั้นเองเป็นโปรโตไทป์เท่านั้น ดังนั้นวิทยานิพนธ์ฉบับนี้จึงนำเทคนิค "Singleton" เข้ามาเพื่อเพิ่มประสิทธิภาพการทำงานในส่วนของกระบวนการ MCS ซึ่งทำให้การคำนวณความครอบคลุมของกระบวนการ MCS มีความถูกต้องมากยิ่งขึ้น โดยจะทำการเลือก Singleton เป็นโปรโตไทป์ก่อนกระบวนการ MCS ดังแสดงในรูปที่ 4.11



รูปที่ 4.11 แผนผังการทำงานของ HGAs ที่ใช้เทคนิค "Singleton"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การเพิ่มประสิทธิภาพการทำงานด้วยเทคนิค "Diversity"

ตามที่ได้อธิบายไปในหัวข้อที่ 3.2 ว่าการทำงานของ GAs เลียนแบบมาจากระบบวิวัฒนาการทางธรรมชาติสำหรับพันธุกรรมของสิ่งมีชีวิตของ "Charles Darwin" ซึ่งกล่าวไว้ว่า "วัฏจักรของพันธุกรรมเกิดมาจากความหลากหลายทางพันธุกรรมที่ได้รับการคัดสรรทางธรรมชาติ" ดังนั้นถ้าประชากรมีความคล้ายคลึงกันมากเกินไป ก็จะทำให้การพัฒนาการเป็นไปได้ช้า หรือไม่มีการพัฒนาการเลย

ซึ่งก็สอดคล้องกับผลการทำงานของ HGAs เนื่องจากพบว่าเมื่อนำวิธี MCS มาทำงานร่วมกับ GAs แม้ว่าจะช่วยให้โครโมโซมที่ได้พัฒนาไปในทิศทางที่ถูกต้อง แต่ก็ทำให้โครโมโซมในประชากรของแต่ละรอบการทำงานมีลักษณะคล้ายคลึงกันด้วย เปรียบเสมือนการค้นหาคำตอบอยู่เพียงบริเวณเดียวบนปริภูมิคำตอบทั้งหมด ดังนั้นเพื่อทำให้การค้นหาคำตอบกว้างขึ้น จึงต้องทำการเพิ่มความหลากหลายให้กับประชากร

วิทยานิพนธ์ฉบับนี้จึงนำเทคนิค "Diversity หรือ ความหลากหลาย" เข้ามาช่วยเพิ่มประสิทธิภาพการทำงานในส่วนของ GAs โดยเทคนิค "ความหลากหลาย" เป็นเทคนิคที่ใช้กันอย่างแพร่หลายในการเพิ่มประสิทธิภาพของ GAs เช่น [16] และ [17] เป็นต้น

ความหลากหลายในวิทยานิพนธ์ฉบับนี้จะถูกคำนวณโดยฟังก์ชั่นคำนวณระยะทางออกมาในรูปของระยะทาง กล่าวคือถ้าออบเจ็คใดๆ มีระยะห่างกันมาก ก็จะหมายความว่ามีความหลากหลายแตกต่างกันมาก ในทางกลับกันถ้าออบเจ็คใดๆ มีระยะห่างกันน้อย ก็จะหมายความว่ามีความหลากหลายแตกต่างกันน้อย

ฟังก์ชั่นคำนวณระยะทางที่ใช้ในการคำนวณค่าความหลากหลายจะไม่ใช้การวัดระยะทางแบบยูคลิดเหมือนกับการการคำนวณความเหมือน เพราะการวัดระยะทางแบบยูคลิดเหมาะกับข้อมูลที่มีค่าคุณลักษณะเฉพาะที่เป็นจำนวนจริง แต่โครโมโซมที่ต้องการคำนวณความหลากหลายนั้นอยู่ในรูป Binary String จึงไม่เหมาะที่จะใช้การวัดระยะทางแบบยูคลิดในการคำนวณ แต่จะใช้การวัดระยะทางแบบแฮมมิ่งแทน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.1 การวัดระยะทางแบบฮัมมิ่ง (Hamming Distance)

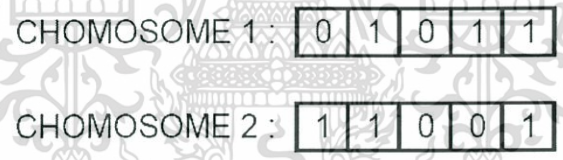
การวัดระยะแบบฮัมมิ่ง [1] เป็นการคำนวณระยะห่างระหว่างวัตถุแบบหนึ่ง โดยในวิทยานิพนธ์ฉบับนี้จะใช้การวัดระยะแบบฮัมมิ่งเป็นฟังก์ชั่นคำนวณระยะทาง เป็นเครื่องมือในการวัดความหลากหลายระหว่างออบเจ็ค สามารถอธิบายให้อยู่ในเชิงคณิตศาสตร์ได้ดังนี้

$$d(x,y) = \sum_{k=1}^{n} x_k - y_k \tag{4.1}$$

โดยกำหนดให้

- $d(x,y)$ คือ ความแตกต่างระหว่างโครโมโซมในประชากร
- x_k และ y_k คือ ค่าลักษณะเฉพาะของออบเจ็ค x และออบเจ็ค y ตามลำดับ
- k คือ จำนวนค่าลักษณะเฉพาะที่ไปได้ทั้งหมด ในที่นี้แทนด้วยจำนวนจริงขนาด n มิติ

การวัดระยะแบบฮัมมิ่งสามารถอธิบายได้ง่ายๆ คือ ผลรวมของจำนวนบิตที่แตกต่างกันระหว่าง 2 ออบเจ็ค ตัวอย่างเช่น กำหนดให้ประชากรตัวอย่างที่ 1 มีสมาชิกเป็นโครโมโซมที่เข้ารหัสอยู่ในรูปของ Binary String จำนวน 2 ตัว (แต่ละโครโมโซมมีขนาด 5 บิต) ดังแสดงในรูปที่ 4.12



รูปที่ 4.12 ประชากรตัวอย่างที่ 1

จากรูปที่ 4.12 แสดงตัวอย่างประชากรตัวอย่างที่ 1 ซึ่งมีสมาชิกเป็นโครโมโซมจำนวน 2 ตัว ซึ่งจะเห็นว่าระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2 มีจำนวนบิตที่แตกต่างกันอยู่ 2 บิต คือ บิตที่ 1 กับ บิตที่ 4 ดังนั้นการวัดระยะแบบฮัมมิ่งระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2 จะเท่ากับ “2”

4.4.2 การประยุกต์ใช้หลักการ “Diversity”

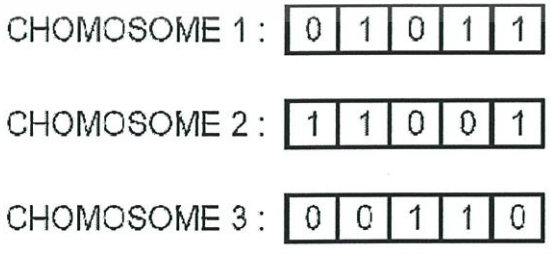
ในวิทยานิพนธ์ฉบับนี้จะนำหลักการ “Diversity” เข้ามาใช้กับ 2 กระบวนการในส่วนของ GAs คือ กระบวนการครอสโอเวอร์ และกระบวนการเลือกประชากรรุ่นถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2.1 การประยุกต์หลักการ "Diversity" ในกระบวนการครอสโอเวอร์ (Diversity Crossover)

กระบวนการครอสโอเวอร์ของ HGAs ตามที่ได้อธิบายในหัวข้อ 4.2.4 นั้น จะทำการสุ่มเลือกโครโมโซมจากกลุ่มประชากรมาใช้เป็นพ่อ-แม่พันธุ์ โดยโอกาสที่โครโมโซมใดๆ จะถูกเลือกจะเป็นอัตราส่วนแปรผันตรงกับค่าความเหมาะสมของโครโมโซมนั้นๆ เพียงอย่างเดียว ทำให้โครโมโซมที่มีค่าความเหมาะสมมากก็จะมีความน่าจะเป็นที่ถูกสุ่มเลือกได้สูง ซึ่งแม้ว่าจะทำให้ได้โครโมโซมที่มีดีมาเป็นพ่อ-แม่พันธ์ แต่โครโมโซมพ่อ-แม่ที่สุ่มได้ อาจจะมีลักษณะคล้ายคลึงกันมาก เพราะโครโมโซมเหล่านี้เป็นผลลัพธ์มาจากกระบวนการ MCS ที่อาศัยหลัก "ความครอบคลุม" ในการเลือกโปรโตไทป์ ทำให้ออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงมีแนวโน้มที่จะถูกเลือกเสมอๆ เป็นผลให้โปรโตไทป์ผลลัพธ์ที่ได้จะมีลักษณะคล้ายคลึงกันส่งผลให้การเฟ้นสุ่มหาคำตอบของ GAs ค้นหาคำตอบอยู่เพียงบริเวณเดียวของปริภูมิคำตอบ ไม่มีการค้นหาคำตอบบริเวณอื่นที่อาจจะมีคำตอบที่ดีกว่า

เทคนิค "Diversity" จึงถูกนำมาใช้เพื่อช่วยเพิ่มความหลากหลายให้กับประชากร โดยจะทำให้โครโมโซมพ่อ-แม่ มีลักษณะไม่คล้ายคลึงกันมากเกินไป ซึ่งจะทำให้โครโมโซมลูกที่ได้มีความหลากหลายตามไปด้วย (เรียกวิธีนี้ว่า "Diversity Crossover") ในวิทยานิพนธ์ฉบับนี้จะใช้การวัดระยะแบบฮัมมิ่งในการวัดความหลากหลายของโครโมโซม โดยจะทำการหาค่า "Diversity Threshold" ซึ่งเป็นค่าเฉลี่ยของความหลากหลายของโครโมโซมทั้งหมดในประชากร จากนั้นใช้ค่า Diversity Threshold เป็นเกณฑ์ของโครโมโซมพ่อ-แม่ที่สุ่มได้ ถ้าโครโมโซมพ่อมีความแตกต่างกับโครโมโซมแม่น้อยกว่าเกณฑ์ที่กำหนด ก็หมายความว่าโครโมโซมพ่อกับโครโมโซมแม่มีความคล้ายคลึงกันมากเกินไป ไม่ควรนำมาเป็นพ่อ-แม่พันธุ์ จึงทำการสุ่มเลือกโครโมโซมพ่อ-แม่ใหม่จนกว่าจะได้โครโมโซมพ่อ-แม่ มีความแตกต่างกันมากกว่าเกณฑ์ที่กำหนด ตัวอย่างเช่น กำหนดให้ประชากรตัวอย่างที่ 2 มีสมาชิกเป็นโครโมโซมที่เข้ารหัสอยู่ในรูปของ Binary String จำนวน 3 ตัว (แต่ละโครโมโซมมีขนาด 5 บิต) ดังแสดงในรูปที่ 4.12



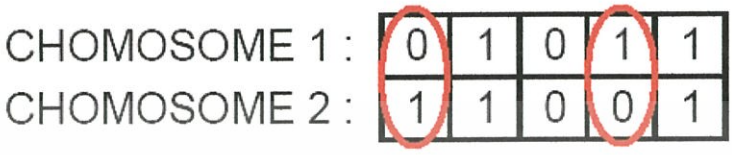
รูปที่ 4.13 ประชากรตัวอย่างที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนของการทำงานมีรายละเอียดดังนี้

- **การคำนวณค่าความหลากหลาย (Diversity Calculation)**

การคำนวณค่าความหลากหลายเป็นการคำนวณความแตกต่างระหว่าง 2 โครโมโซมที่เป็นไปได้ทั้งหมดในประชากรตัวอย่างที่ 2 โดยใช้การวัดระยะทางแบบฮัมมิ่ง ดังแสดงตัวอย่างในรูปที่ 4.14

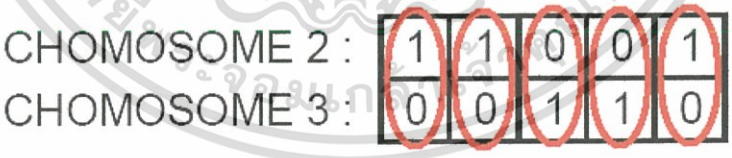


รูปที่ 4.14 ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2

จากรูปที่ 4.14 จะเห็นว่าความหลากหลายระหว่างโครโมโซม 1 กับโครโมโซม 2 เท่ากับ “2” เพราะมีบิตที่แตกต่างกัน 2 บิต ได้แก่ บิตที่ 1 และบิตที่ 4

รูปที่ 4.15 ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 3

จากรูปที่ 4.15 จะเห็นว่าความหลากหลายระหว่างโครโมโซม 1 กับโครโมโซม 3 เท่ากับ “3” เพราะมีบิตที่แตกต่างกัน 3 บิต ได้แก่ บิตที่ 2, บิตที่ 3 และบิตที่ 5



รูปที่ 4.16 ค่าความหลากหลายระหว่าง CHROMOSOME 2 กับ CHROMOSOME 3

จากรูปที่ 4.16 จะเห็นว่าความหลากหลายระหว่างโครโมโซม 2 กับโครโมโซม 3 เท่ากับ “5” เพราะมีบิตที่แตกต่างกัน 5 บิต ได้แก่ บิตที่ 1, บิตที่ 2, บิตที่ 3, บิตที่ 4 และบิตที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **การคำนวณค่า Diversity Threshold (Diversity Threshold Calculation)**

การคำนวณค่า Diversity Threshold เป็นการคำนวณค่าเฉลี่ยความหลากหลายของโครโมโซมทั้งหมดในประชากร ซึ่งทำได้โดยการคำนวณค่าความหลากหลายรวมของทุกโครโมโซมในประชากร จากนั้นนำมาหารด้วยจำนวนโครโมโซมทั้งหมดในประชากร สามารถอธิบายให้อยู่ในเชิงคณิตศาสตร์ได้ดังนี้

$$\overline{d(x)} = \frac{\sum_{k=1}^{n} x_k}{n} \tag{4.2}$$

โดยกำหนดให้

- $\overline{d(x)}$ คือ ค่าเฉลี่ยความหลากหลายของประชากรทั้งหมด
- x_k คือ ค่าความหลากหลายของโครโมโซมตัวที่ k
- k คือ จำนวนโครโมโซมทั้งหมดในประชากร แทนด้วยจำนวนจริงจำนวน n ตัว

จากประชากรตัวอย่างที่ 2ซึ่งมีสมาชิกเป็นโครโมโซมจำนวน 3 ตัว สามารถคำนวณค่าเฉลี่ยความหลากหลายของประชากรตัวอย่างที่ 2 ทั้งหมดได้ดังนี้

- ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2 เท่ากับ “2”
- ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 3 เท่ากับ “3”
- ค่าความหลากหลายระหว่าง CHROMOSOME 2 กับ CHROMOSOME 3 เท่ากับ “5”

ดังนั้นค่าเฉลี่ยความหลากหลายของโครโมโซมทั้งหมดประชากรตัวอย่างที่ 2 (Diversity Threshold) เท่ากับ (2 + 3 + 5) / 3 = “3.34”

หมายความว่าถ้าโครโมโซมพ่อ-แม่ ที่สุ่มเลือกได้มีความแตกต่างกันน้อยกว่า Diversity Threshold ซึ่งเท่ากับ “3.34” ก็จะทำการสุ่มเลือกใหม่ และจะทำการสุ่มเลือกจนกว่าจะได้โครโมโซมพ่อ-แม่ที่มีความแตกต่างกันมากกว่า “3.34” ตัวอย่างเช่น ถ้าสุ่มเลือกได้ CHROMOSOME 1 กับ CHROMOSOME 2 ดังแสดงในรูปที่ 4.16

CHOMOSOME 1 :	0	1	0	1	1
CHOMOSOME 2 :	1	1	0	0	1

รูปที่ 4.17 ตัวอย่างโครโมโซมที่สุ่มเลือกได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.17 แสดงตัวอย่างโครโมโซมที่สุ่มเลือกได้ ซึ่งก็คือ CHROMOSOME 1 กับ CHROMOSOME 2 ซึ่งค่าความหลากหลายของ 2 โครโมโซมนี้เท่ากับ "2" เพราะมีบิตที่แตกต่างกัน 2 บิต ซึ่งจะเห็นว่ามีค่าน้อยกว่า Diversity Threshold ที่มีค่าเท่ากับ "3.34" ดังนั้นจะทำการสุ่มโครโมโซมพ่อ-แม่ใหม่ ตัวอย่างเช่น ถ้าสุ่มเลือกได้ CHROMOSOME 2 กับ CHROMOSOME 3 ดังแสดงในรูปที่ 4.18

CHOMOSOME 2 :	1	1	0	0	1
CHOMOSOME 3 :	0	0	1	1	0

รูปที่ 4.18 ตัวอย่างโครโมโซมที่สุ่มเลือกใหม่

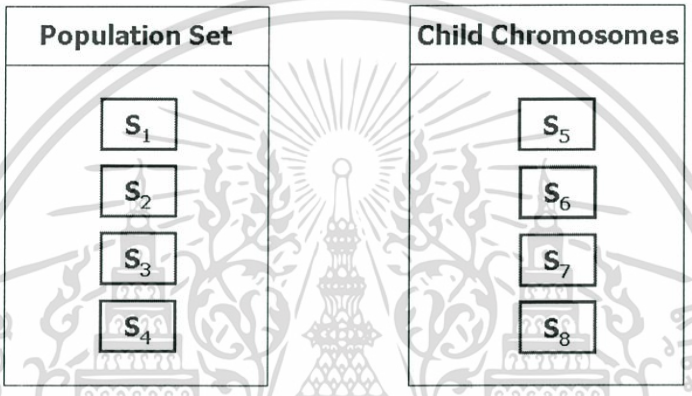
จากรูปที่ 4.18 แสดงตัวอย่างโครโมโซมที่สุ่มเลือกใหม่ ก็คือ CHROMOSOME 2 กับ CHROMOSOME 3 ซึ่งค่าความหลากหลายความระหว่าง 2 โครโมโซมนี้เท่ากับ "5" เพราะมีบิตที่แตกต่างกัน 5 บิต ซึ่งจะเห็นว่ามีค่ามากกว่า Diversity Threshold ซึ่งมีค่าเท่ากับ "3.34" ดังนั้นโครโมโซมนี้จึงเหมาะสมที่จะใช้เป็นโครโมโซมพ่อ-แม่

4.4.2.2 การประยุกต์หลักการ "Diversity" ในกระบวนการเลือกสรรประชากรรุ่นถัดไป (Diversity Selection)

กระบวนการเลือกประชากรรุ่นถัดไปของ HGAs ตามที่ได้อธิบายในหัวข้อ 4.2.6 นั้น จะทำการเลือกเก็บโครโมโซมบางส่วนจากโครโมโซมที่มีอยู่ทั้งหมดไปเป็นประชากรโครโมโซมในรุ่นถัดไป โดยในการเลือกประชากรจะเลือกตามโครโมโซมที่มีค่าความเหมาะสมมากที่สุดก่อน และจะทำการเลือกโครโมโซมจำนวน N_{pop} ตัว ไปเป็นประชากรในรอบถัดไป ซึ่งแม้ว่าโครโมโซมที่มีค่าความเหมาะสมมากจะเป็นโครโมโซมที่ดี และมีแนวโน้มที่จะพัฒนาไปเป็นโครโมโซมที่ดีขึ้น แต่ในที่นี้โครโมโซมที่เลือกได้ไม่ได้นำไปเป็นประชากรฝึกสอนโดยตรง แต่จะนำโครโมโซมเหล่านี้ไปผ่านกระบวนการ MCS ก่อน ซึ่งถ้าโครโมโซมเหล่านี้มีความคล้ายคลึงกันมากเกินไป ก็จะทำให้โปรโตไทป์ผลลัพธ์มีความคล้ายคลึงกันตามไปด้วย จะเห็นว่าเป็นการสิ้นเปลืองโดยเปล่าประโยชน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เช่นเดียวกันกับกระบวนการครอสโอเวอร์ที่เทคนิค “Diversity” ถูกนำใช้มาเพื่อช่วยเพิ่มความหลากหลาย ทำให้โครโมโซมที่เลือกไปเป็นประชากรรุ่นถัดไปมีลักษณะที่ไม่คล้ายกันมากจนเกินไป (เรียกวิธีนี้ว่า “Diversity Selection”) และเช่นกันวิทยานิพนธ์ฉบับนี้จะใช้การวัดระยะแบบฮัมมิ่งในการวัดความหลากหลายของโครโมโซม โดยจะทำการเลือกโครโมโซมไปเป็นประชากรในรุ่นถัดไปโดยพิจารณาจากค่าความเหมาะสมจำนวน $N_{pop}/2$ ตัว ส่วนอีก $N_{pop}/2$ ตัว จะทำการเลือกโครโมโซมโดยพิจารณาจากค่าความหลากหลาย ซึ่งเป็นค่าเฉลี่ยความหลากหลายของแต่ละโครโมโซม ตัวอย่างเช่น กำหนดให้ประชากรมีสมาชิกเป็นโครโมโซมจำนวน 4 ตัว และมีโครโมโซมลูกจำนวน 4 ตัว ดังแสดงในรูปที่ 4.19



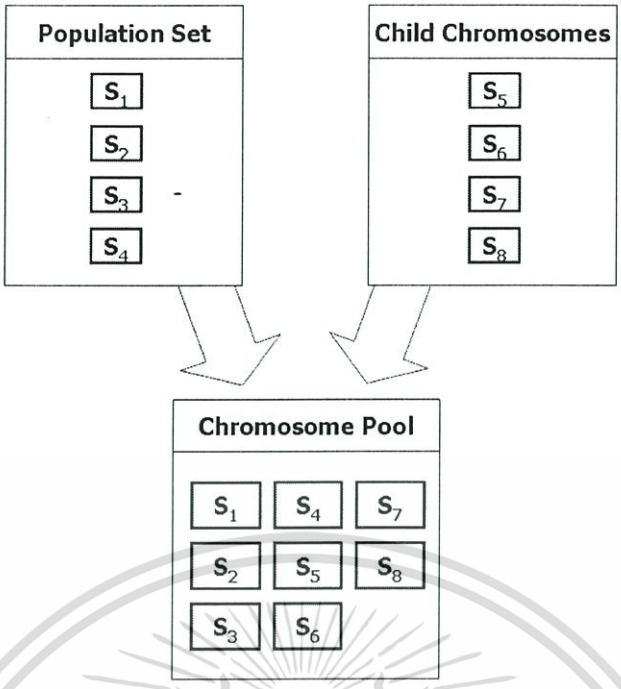
รูปที่ 4.19 ตัวอย่างโครโมโซมในประชากร และโครโมโซมลูก

ขั้นตอนการทำงานมีรายละเอียดดังนี้

- **การนำโครโมโซมทั้งหมดมารวมกัน (Chromosome Pooling)**

การนำโครโมโซมทั้งหมดมารวมกันเป็นการกระบวนการนำโครโมโซมในประชากรกับโครโมโซมลูกที่ได้จากกระบวนการครอสโอเวอร์ และกระบวนการกลายพันธุ์แล้ว มารวมกันไว้เป็นกลุ่มโครโมโซมร่วม ดังแสดงตัวอย่างในรูปที่ 4.19

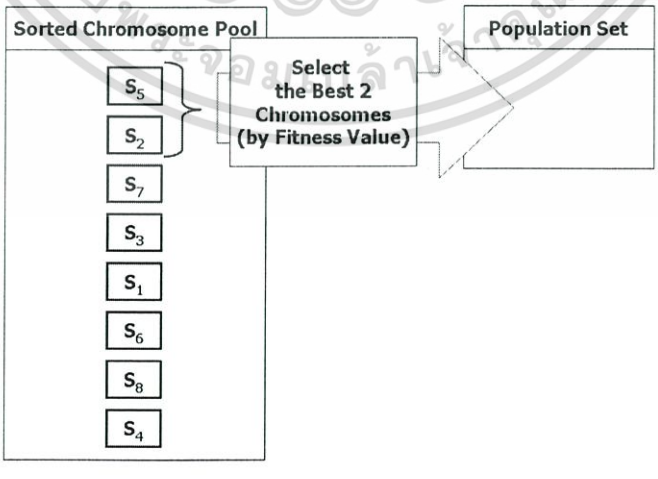
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 ตัวอย่างการสร้างกลุ่มโครโมโซมร่วม

- **การเลือกโครโมโซมด้วยค่าความเหมาะสม (Fitness Selection)**

การเลือกโครโมโซมด้วยค่าความเหมาะสมเป็นการนำโครโมโซมในกลุ่มโครโมโซมร่วมมาคำนวณหาค่าความเหมาะสมโดยใช้ฟังก์ชั่นคำนวณค่าความเหมาะสม จากนั้นทำการเรียงลำดับโครโมโซมทั้งหมดในกลุ่มโครโมโซมร่วมตามค่าความเหมาะสมจากมากสุดไปน้อยสุด และทำการเลือกโครโมโซมที่มีค่าความเหมาะสมมากที่สุดจำนวน $N_{pop}/2$ ตัว ไปเป็นประชากรรุ่นถัดไป ดังแสดงตัวอย่างในรูปที่ 4.21



รูปที่ 4.21 ตัวอย่างการเลือกโครโมโซมโดยใช้ค่าความเหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.21 แสดงตัวอย่างการเลือกโครโมโซมด้วยค่าความเหมาะสมจากประชากรที่มีสมาชิกเป็นโครโมโซมจำนวน 4 ตัว ดังนั้นจะทำการเลือกโครโมโซมด้วยค่าความเหมาะสมจำนวน 2 ตัว โดยในที่นี้สมมติว่าโครโมโซม S_5 และโครโมโซม S_2 มีค่าความเหมาะสมมากที่สุด จึงทำการเลือกโครโมโซม 2 ตัวนี้ไปเป็นประชากรรุ่นถัดไป

- **การเลือกโครโมโซมด้วยค่าความหลากหลาย (Diversity Selection)**

การเลือกโครโมโซมด้วยค่าความหลากหลายเป็นการนำโครโมโซมในกลุ่มโครโมโซมร่วมที่เหลือจากกระบวนการเลือกโครโมโซมด้วยค่าความเหมาะสมมาคำนวณหาค่าความหลากหลาย ที่คำนวณโดยใช้การวัดระยะทางแบบฮัมมิ่ง จากนั้นทำการเรียงลำดับโครโมโซมทั้งหมดในกลุ่มโครโมโซมร่วมตามค่าความหลากหลาย จากมากสุดไปน้อยสุด และทำการเลือกโครโมโซมที่มีค่าความหลากหลายมากที่สุดจำนวน $N_{pop}/2$ ตัว ไปเป็นประชากรรุ่นถัดไป ดังแสดงตัวอย่างในรูปที่ 4.22

ตัวอย่างเช่น จากประชากรตัวอย่างที่ 2 ที่ มีสมาชิกเป็นโครโมโซมจำนวน 3 ตัว จะสามารถคำนวณค่าความหลากหลายได้ดังนี้

ค่าความหลากหลายของ CHROMOSOME 1 สามารถคำนวณได้โดย

- ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2 ซึ่งเท่ากับ 2

รวมกับ

- ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 3 ซึ่งเท่ากับ 3

ดังนั้นค่าความหลากหลายรวมของ CHROMOSOME 1 คือ 2+3 = “5”

ค่าความหลากหลายของ CHROMOSOME 2 สามารถคำนวณได้โดย

- ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 2 ซึ่งเท่ากับ 2

รวมกับ

- ค่าความหลากหลายระหว่าง CHROMOSOME 2 กับ CHROMOSOME 3 ซึ่งเท่ากับ 5

ดังนั้นค่าความหลากหลายรวมของ CHROMOSOME 2 คือ 2+5 = “7”

ค่าความหลากหลายของ CHROMOSOME 3 สามารถคำนวณได้โดย

- ค่าความหลากหลายระหว่าง CHROMOSOME 1 กับ CHROMOSOME 3 ซึ่งเท่ากับ 3

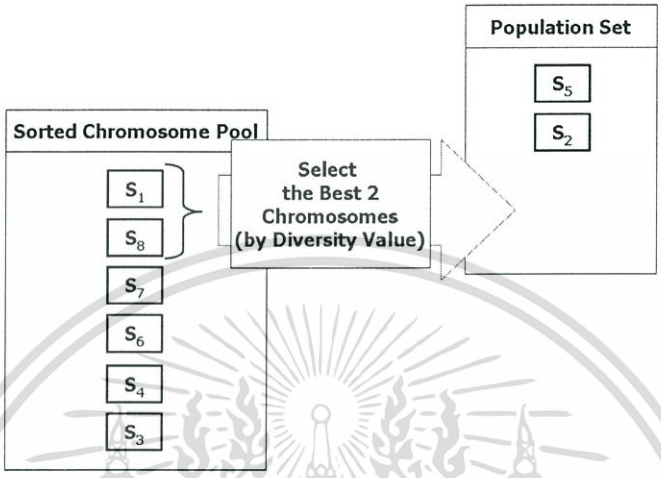
รวมกับ

- ค่าความหลากหลายระหว่าง CHROMOSOME 2 กับ CHROMOSOME 3 ซึ่งเท่ากับ 5

ดังนั้นค่าความหลากหลายรวมของ CHROMOSOME 3 คือ 3+5 = “8”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นถ้าทำการเลือกโครโมโซมด้วยค่าความหลากหลายจากตัวอย่างประชากรตัวอย่างที่ 2 ก็ต้องทำการเลือก CHROMOSOME 3 ก่อนเพราะ CHROMOSOME 3 มีค่าความหลากหลายรวมเท่ากับ "8" จากนั้นจึงเลือก CHROMOSOME 2 และ CHROMOSOME 1 ตามลำดับ



รูปที่ 4.22 ตัวอย่างการเลือกโครโมโซมโดยใช้ค่าความหลากหลาย

จากรูปที่ 4.22 แสดงแสดงตัวอย่างการเลือกโครโมโซมโดยใช้ค่าความหลากหลายจากประชากรที่มีสมาชิกเป็นโครโมโซมจำนวน 4 ตัว ดังนั้นจะทำการเลือกโครโมโซมด้วยค่าความหลากหลายจำนวน 2 ตัว โดยสมมติว่าโครโมโซม S_6 และโครโมโซม S_3 มีค่าความหลากหลายมากที่สุด จึงทำการเลือกโครโมโซม 2 ตัวนี้ไปรวมกับโครโมโซมที่เลือกด้วยค่าความเหมาะสม ที่เลือกแล้วอีก 2 ตัวเพื่อเป็นประชากรรุ่นถัดไป

4.5 HGAs แบบต่างๆ

ตามที่ได้อธิบายไปข้างต้นว่า HGAs เป็นการทำงานร่วมกันของวิธี MCS กับ GAs ซึ่งเมื่อนำทั้ง 2 วิธี มาทำงานร่วมกันจำเป็นต้องมีการปรับแต่งรายละเอียดขั้นตอนการทำงานบางอย่างเพื่อให้ทั้ง 2 วิธี ทำงานเข้ากันอย่างมีประสิทธิภาพมากที่สุด อีกทั้งยังมีการนำเทคนิค "Singleton" และเทคนิค "Diversity" เข้ามาเพิ่มประสิทธิภาพให้กับ HGAs อีกด้วย ดังนั้นในหัวข้อนี้จะแสดง HGAs แบบต่างๆ เพื่อแสดงให้เห็นว่า HGAs แบบที่ได้นำเสนอไปมีประสิทธิภาพในการทำงานสูงสุดเมื่อเทียบกับ HGAs แบบอื่นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

HGAs แบ่งได้ 3 แบบตามเทคนิคที่ใช้ ดังนี้

- HGAs แบบของเทคนิคการเลือกของกระบวนการ MCS แบ่งเป็น
  - HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย
  - HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมด
- HGAs แบบของเทคนิค “Singleton” แบ่งเป็น
  - HGAs แบบที่ใช้เทคนิค “Singleton”
  - HGAs แบบที่ไม่ใช้เทคนิค “Singleton”
- HGAs แบบของเทคนิค “Diversity” แบ่งเป็น
  - HGAs แบบที่ใช้เทคนิค “Diversity Crossover” และ “Diversity Selection”
  - HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Crossover”
  - HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Selection”
  - HGAs แบบที่ไม่ใช้เทคนิค “Diversity Crossover” และ “Diversity Selection”

4.5.1 HGAs แบบของเทคนิคการเลือกของกระบวนการ MCS

HGAs แบบของเทคนิคการเลือกของกระบวนการ MCS เป็น HGAs แบบที่ใช้กระบวนการ MCS ที่แตกต่างกัน กล่าวคือแบบหนึ่งใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย ส่วนอีกแบบกระบวนการ MCS แบบที่เลือกจากทั้งหมด

HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย มีขั้นตอนการทำงานอ้างอิงตามหัวข้อที่ 4.2 กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค “Singleton”, เทคนิค “Diversity Crossover” และเทคนิค “Diversity Selection” เข้ามาทำงานร่วมด้วย

ส่วน HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมด ก็จะมีขั้นตอนการทำงานตามหัวข้อที่ 4.2 เช่นกัน แต่จะแตกต่างกันตรงที่จะใช้ประบวนการ MCS แบบที่เลือกโปรโตไทป์จากชุดข้อมูลฝึกสอนทั้งหมด กล่าวคือโปรโตไทป์ผลลัพธ์ที่ได้อาจจะไม่เป็นเซตย่อยของ Candidate Set ก็ได้ และมีการนำเทคนิค “Singleton”, เทคนิค “Diversity Crossover” และเทคนิค “Diversity Selection” เข้ามาทำงานร่วมด้วยเช่นกัน

ตัวอย่างเช่น จากตัวอย่างชุดข้อมูลที่ 2 และ Candidate Set ตัวอย่างที่ 2 คือ {A3, B1} เมื่อทำการคำนวณค่าความสามารถในการครอบคลุม ก็จะได้ตารางค่าความสามารถในการครอบคลุมดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2

Candidate Set = {A3, B1}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	0	0	0	0	0
A3	1	0	1	0	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	1	1	1
B2	0	0	0	0	1	1	1
B3	0	0	0	0	0	1	1
CV	3	2	2	1	2	3	3

จากตารางที่ 4.4 จะเห็นว่ากระบวนการ MCS แบบที่เลือกจากเซตย่อยจะต้องเลือกออบเจ็ค A3 หรือออบเจ็ค B1 เป็นโปรโตไทป์ เพราะออบเจ็ค A3 และออบเจ็ค B1 อยู่ใน Candidate Set และมีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ "2"

การเลือกจากเซตย่อยจะทำให้โปรโตไทป์ผลลัพธ์จะถูกจำกัดให้เป็นเซตย่อยของ Candidate Set เสมอ ดังที่แสดงไปในหัวข้อที่ 4.2.3.2 ซึ่งจะเห็นว่าถ้าใช้กระบวนการ MCS แบบที่จากเซตย่อยจะได้โปรโตไทป์ผลลัพธ์ = {A3, B1} เพราะถูกจำกัดโดย Candidate Set

ดังนั้นจึงทำการทดลองใช้กระบวน MCS แบบที่เลือกจากทั้งหมด ซึ่งจะทำให้โปรโตไทป์ผลลัพธ์ไม่ถูกจำกัดให้เป็นเซตย่อยของ Candidate Set

จากตารางที่ 4.4 ถ้าใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมด ก็จะต้องทำการเลือกออบเจ็ค A1 หรือออบเจ็ค B2 หรือออบเจ็ค B3 เป็นโปรโตไทป์ เพราะออบเจ็ค A1, ออบเจ็ค B2 และออบเจ็ค B3 มีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ "3" ในที่นี้สมมติว่าเลือกว่าออบเจ็ค B2 เป็นโปรโตไทป์ จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค B1, ออบเจ็ค B2 และออบเจ็ค B3ที่ถูกครอบคลุมโดยออบเจ็ค B2 แล้วออก ดังผลลัพธ์ในตารางที่ 4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B2 เป็นโปรโตไทป์แล้ว

Candidate Set = {A3, B1}

Prototype Set = {B2}

	A1	A2	A3	A4	B1	B2	B3
A1	1	1	1	0	0	0	0
A2	1	1	0	0	0	0	0
A3	1	0	1	0	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	X	X	X
B2	0	0	0	0	X	X	X
B3	0	0	0	0	0	X	X
CV	3	2	2	1	0	0	0

จากนั้นทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดลำดับต่อไปเป็นโปรโตไทป์ ซึ่งจากตารางที่ 4.5 จะเห็นว่าออบเจ็ค A1 มีค่าความสามารถในการครอบคลุมสูงสุดเท่ากับ “3” ดังนั้นจึงทำการเลือกออบเจ็ค A1 เป็นโปรโตไทป์ จากนั้นทำการคำนวณค่าความสามารถในการครอบคลุมของแต่ละออบเจ็คใหม่ โดยทำการลดทอนค่าความสามารถในการครอบคลุมของออบเจ็ค A1, ออบเจ็ค A2 และออบเจ็ค A3 ที่ถูกครอบคลุมโดยออบเจ็ค A1 แล้วออก ดังแสดงในตารางที่ 4.6

ตารางที่ 4.6 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B2 และ A1 เป็นโปรโตไทป์แล้ว

Candidate Set = {A3, B1}

Prototype Set = {A1, B2}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	0	0	0	0	0
A3	X	0	X	0	0	0	0
A4	0	0	0	1	0	0	0
B1	0	0	0	0	X	X	X
B2	0	0	0	0	X	X	X
B3	0	0	0	0	0	X	X
CV	0	0	0	1	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในขั้นตอนสุดท้ายก็ทำการเลือกออบเจ็ค A4 ที่เหลืออยู่เพียงตัวเดียวเป็นโปรโตไทป์ และเมื่อทำการเลือกออบเจ็ค A4 เป็นโปรโตไทป์แล้ว ค่าความสามารถในการครอบคลุมของออบเจ็ค A4 ก็จะถูกลดทอนออก ดังแสดงในตารางที่ 4.7

ตารางที่ 4.7 ค่าความสามารถในการครอบคลุมของตัวอย่างชุดข้อมูลที่ 2 โดยใช้ Candidate Set ตัวอย่างที่ 2 หลังเลือกออบเจ็ค B2, A1 และ A4 เป็นโปรโตไทป์แล้ว

Candidate Set = {A3, B1}

Prototype Set = {A1, A4, B2}

	A1	A2	A3	A4	B1	B2	B3
A1	X	X	X	0	0	0	0
A2	X	X	0	0	0	0	0
A3	X	0	X	0	0	0	0
A4	0	0	0	X	0	0	0
B1	0	0	0	0	X	X	X
B2	0	0	0	0	X	X	X
B3	0	0	0	0	0	X	X
CV	0	0	0	0	0	0	0

เมื่อพิจารณาตารางที่ 4.7 จะเห็นว่าค่าความสามารถในการครอบคลุมของทุกออบเจ็คเท่ากับ "0" เพราะทุกออบเจ็คในชุดข้อมูลที่ 2 ถูกครอบคลุมโดยโปรโตไทป์ที่เลือกแล้ว และเมื่อทุกออบเจ็คถูกครอบคลุมหมดแล้วก็เป็นอันสิ้นสุดการทำงานของกระบวนการ MCS แบบที่เลือกจากทั้งหมด สรุปจะได้โปรโตไทป์ผลลัพธ์ คือ {A1, A4, B2} ซึ่งเป็นโปรโตไทป์ที่มีความสอดคล้องกับชุดข้อมูลฝึกสอน (เพราะสามารถครอบคลุมออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนได้)

จากตัวอย่างข้างต้นจะเห็นว่ากระบวนการ MCS แบบที่เลือกจากทั้งหมดให้ผลลัพธ์ที่ดีกว่ากระบวนการ MCS แบบที่เลือกจากเซตย่อย เพราะให้ผลลัพธ์ที่สอดคล้องกับชุดข้อมูลฝึกสอน จึงตั้งสมมติฐานว่าการเลือกจากทั้งหมดช่วยทำให้ผลลัพธ์ที่ดีกว่าการเลือกจากเซตย่อย ดังนั้นจำนวนโปรโตไทป์ผลลัพธ์ของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อยกับจำนวนโปรโตไทป์ผลลัพธ์ของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมดจึงถูกนำมาเปรียบเทียบกัน (รายละเอียดแสดงในภาคผนวก จ.)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรายละเอียดผลการทดลองที่แสดงในภาคผนวก จ. พบว่า แม้ตามทฤษฎีกระบวนการ MCS แบบที่เลือกจากทั้งหมดให้ผลลัพธ์ที่ดีกว่ากระบวนการ MCS แบบที่เลือกจากเซตย่อย แต่เมื่อพิจารณาจากผลการทดลองจะเห็นว่ากระบวนการ MCS แบบที่เลือกจากเซตย่อยให้ผลลัพธ์ที่ดีกว่ากระบวนการ MCS แบบที่เลือกจากทั้งหมด เนื่องมาจากการคำนวณความครอบคลุมที่ผิดพลาด เพราะแม้ว่าจะทำการเลือกออบเจ็คจากชุดข้อมูลฝึกสอน แต่ในการคำนวณความครอบคลุมก็ยังใช้ Candidate Set ในการคำนวณอยู่ดี ดังนั้นเมื่อมีการเลือกออบเจ็คนอกเหนือ Candidate Set เป็นโปรโตไทป์ก็จะทำให้ความครอบคลุมที่คำนวณไว้ผิดพลาดไป

ดังนั้นวิทยานิพนธ์ฉบับนี้จึงนำเสนอ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อยในการทำงาน แต่ทั้งนี้ทั้งนั้น Candidate Set ที่ใช้ในการทำงานเริ่มต้นต้องมีจำนวนข้อมูลไม่น้อยจนเกินไป ซึ่งสามารถทำได้โดยกำหนดค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) ให้มีค่ามากๆ

4.5.2 HGAs แบบของเทคนิค "Singleton"

HGAs แบบของเทคนิค "Singleton" เป็น HGAs แบบที่ใช้เทคนิค "Singleton" ที่แตกต่างกัน กล่าวคือแบบหนึ่งจะใช้เทคนิค "Singleton" ส่วนอีกแบบจะไม่ใช้เทคนิค "Singleton"

ขั้นตอนการทำงานของ HGAs แบบที่ใช้เทคนิค "Singleton" จะอ้างอิงตามหัวข้อที่ 4.2 กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton", เทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วย

ส่วน HGAs แบบที่ไม่ใช้เทคนิค "Singleton" จะอ้างอิงตามหัวข้อที่ 4.2 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำเทคนิค "Singleton" เข้ามาทำงานร่วมด้วย กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วยเท่านั้น

ตามที่ได้อธิบายเกี่ยวกับเทคนิค "Singleton" ไปในหัวข้อที่ 4.3 ว่าเป็นเทคนิคที่ช่วยทำให้กระบวนการ MCS สามารถเลือกออบเจ็คที่เป็น "Singleton" ซึ่งเป็นออบเจ็คที่ต้องเลือกเป็นโปรโตไทป์เพื่อให้โปรโตไทป์ผลลัพธ์มีความสอดคล้องกับชุดข้อมูลฝึกสอน เปรียบเสมือนการมีกุญแจที่สำคัญในการค้นหาคำตอบทำให้การคำนวณความครอบคลุมมีความถูกต้องมากขึ้นจึงตั้งสมมติฐานว่าเทคนิค "Singleton" ช่วยทำให้ผลลัพธ์ดีขึ้น ดังนั้นจำนวนโปรโตไทป์ของ HGAs แบบที่ใช้เทคนิค "Singleton" กับจำนวนโปรโตไทป์ของ HGAs แบบที่ไม่ใช้เทคนิค "Singleton" จึงถูกนำมาเปรียบเทียบกัน (รายละเอียดแสดงในภาคผนวก ฉ.)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรายละเอียดผลการทดลองที่แสดงในภาคผนวก ฉ. พบว่าเป็นไปตามสมมติฐานที่ตั้งไว้ กล่าวคือในกรณีความสอดคล้อง 100% นั้น HGAs แบบที่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ในทุกชุดข้อมูล

ส่วนในกรณีความสอดคล้อง 99% และความสอดคล้อง 98% จะเห็นว่าในชุดข้อมูลที่มีข้อมูลจำนวนน้อยๆ อย่างเช่น ชุดข้อมูล IRIS จะเห็นว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้เทคนิค “Singleton” เพราะชุดข้อมูล IRIS มีข้อมูลจำนวนไม่มากนัก ดังนั้น HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ยังสามารถเฟ้นสุ่มหาคำตอบที่ดีได้ ประกอบกับเทคนิค “Singleton” นั้นทำการเลือกออบเจ็คที่ไม่มีออบเจ็คอื่นสามารถครอบคลุมได้เป็นโปรโตไทป์ ซึ่งเป็นสิ่งที่ดีสำหรับกรณีความสอดคล้อง 100% แต่สำหรับความสอดคล้อง 99% และความสอดคล้อง 98% ที่โปรโตไทป์ผลลัพธ์ไม่จำเป็นต้องครอบคลุมชุดข้อมูลฝึกสอนทั้งหมดนั้น Singleton อาจจะไม่ใช่สิ่งที่จำเป็น และเมื่อพิจารณาต่อไปก็จะเห็นว่ายิ่งชุดข้อมูลมีข้อมูลจำนวนมากขึ้น HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ก็จะยิ่งให้ผลลัพธ์ที่แย่ลง เพราะในชุดข้อมูลมีข้อมูลจำนวนมาก ก็จะมีปริภูมิคำตอบจำนวนมหาศาล ทำให้ HGAs แบบที่ไม่มีเทคนิค “Singleton” ช่วยไม่สามารถเฟ้นสุ่มหาคำตอบที่ดีได้ และตามที่ได้อธิบายไปข้างต้นว่า ในการเลือกโปรโตไทป์ที่ดีนั้นโปรโตไทป์ผลลัพธ์ควรจะต้องมีความสอดคล้องกับชุดข้อมูลเพื่อยืนยันว่าประสิทธิภาพการจำแนกข้อมูลที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิงมีความใกล้เคียงกันประสิทธิภาพการจำแนกข้อมูลที่มีชุดข้อมูลเป็นข้อมูลอ้างอิง และจากผลการทดลองข้างต้นก็จะเห็นว่าในกรณีความสอดคล้อง 100% นั้น HGAs แบบที่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ในทุกชุดข้อมูล

ดังนั้นวิทยานิพนธ์ฉบับนี้จึงนำเสนอ HGAs แบบที่ใช้เทคนิค “Singleton” ในการทำงาน

4.5.3 HGAs แบบของเทคนิค “Diversity”

HGAs แบบของเทคนิค “Diversity” เป็น HGAs แบบที่ใช้เทคนิค “Diversity” ที่แตกต่างกัน 4 แบบ ดังนี้

1. HGAs แบบที่ใช้ทั้งเทคนิค “Diversity Crossover” และ “Diversity Selection”
2. HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Crossover”
3. HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Selection”
4. HGAs แบบที่ไม่ใช้ทั้งเทคนิค “Diversity Crossover” และ “Diversity Selection”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการทำงานของ HGAs แบบที่ใช้ทั้งเทคนิค “Diversity Crossover” และ “Diversity Selection” จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค “Singleton”, เทคนิค “Diversity Crossover” และเทคนิค “Diversity Selection” เข้ามาทำงานร่วมด้วย

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Crossover” จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำเทคนิค “Diversity Crossover” เข้ามาทำงานร่วมด้วย กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค “Singleton” และเทคนิค “Diversity Crossover” เข้ามาทำงานร่วมด้วยเท่านั้น

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Selection” จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำเทคนิค “Diversity Selection” เข้ามาทำงานร่วมด้วย กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค “Singleton” และเทคนิค “Diversity Selection” เข้ามาทำงานร่วมด้วยเท่านั้น

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ไม่ใช้ทั้งเทคนิค “Diversity Crossover” และ “Diversity Selection” จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำทั้งเทคนิค “Diversity Crossover” และเทคนิค “Diversity Selection” เข้ามาทำงานร่วมด้วย กล่าวคือใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค “Singleton” เข้ามาทำงานร่วมด้วยเท่านั้น

ตามที่ได้อธิบายเกี่ยวกับเทคนิค “Diversity” ไปในหัวข้อที่ 4.4 ว่าเป็นเทคนิคที่ช่วยเพิ่มความหลากหลายให้กับประชากร เนื่องมาจากการนำเทคนิค MCS มาช่วยไกด์ทำให้ GAs สามารถเฟ้นสุ่มหาคำตอบได้ดีขึ้น แต่ก็จะทำให้โครโมโซมมีลักษณะที่คล้ายคลึงกัน ซึ่งจะทำให้ประสิทธิภาพการทำงานของ HGAs ลดลง เปรียบเสมือนการทำงานซ้ำซ้อน ทำการค้นหาคำตอบอยู่เพียงบริเวณเดียวบนปริภูมิคำตอบทั้งหมด ดังนั้นเทคนิค “Diversity” จึงเปรียบเสมือนการขยายบริเวณการค้นหาคำตอบให้กว้างขึ้น มีการเปลี่ยนไปค้นหาคำตอบบริเวณอื่นที่อาจจะมีคำตอบที่ดีกว่า จึงตั้งสมมติฐานว่าเทคนิค “Diversity” ช่วยทำให้ผลลัพธ์ดีขึ้น ดังนั้นจำนวนโปรโตไทป์ของ HGAs แบบที่ใช้เทคนิค “Diversity” กับจำนวนโปรโตไทป์ของ HGAs แบบที่ไม่ใช้เทคนิค “Diversity” จึงถูกนำมาเปรียบเทียบกัน (รายละเอียดแสดงในภาคผนวก ช.)

จากรายละเอียดผลการทดลองที่แสดงในภาคผนวก ช. พบว่าเป็นไปตามสมมติฐานที่ตั้งไว้ กล่าวคือ HGAs แบบที่ใช้เทคนิค “Diversity” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Diversity” ทุกแบบ ในทุกกรณีความสอดคล้อง และในทุกชุดข้อมูล

ดังนั้นวิทยานิพนธ์ฉบับนี้จึงนำเสนอ HGAs แบบที่ใช้เทคนิค “Diversity” ในการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์ของวิธี MCS, GAs และ HGAs (ซึ่งเป็นวิธีที่วิทยานิพนธ์ฉบับนี้นำเสนอ) โดยแบ่งเป็น 2 การทดลอง ดังนี้

1. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม
2. การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จากวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม

โดยหลักในการเลือกโปรโตไทป์ของวิธี MCS นั้นจะทำการพิจารณาเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดเป็นโปรโตไทป์ แต่จากตารางที่ 3.5 ในบทที่ 3 จะเห็นว่ามีโอกาสที่ออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจะมีมากกว่า 1 ออบเจ็ค ซึ่ง "วิธี MCS แบบดั้งเดิม" ก็ไม่ได้ระบุวิธีการพิจารณาเลือกในกรณีนี้ไว้ ซึ่งการจากสังเกตเบื้องต้นพบว่าลำดับการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดที่แตกต่างกันก็จะทำให้ได้โปรโตไทป์ผลลัพธ์ที่แตกต่างกันด้วย ดังนั้นจึงทำการทดลองเพื่อวิเคราะห์ว่าลำดับการเลือกที่แตกต่างกันจะส่งผลอย่างไรกับโปรโตไทป์ผลลัพธ์ แต่การหาโปรโตไทป์ผลลัพธ์ในทุกกรณีที่เป็นไปได้นั้นต้องใช้เวลาในการคำนวณเป็นเวลานาน ดังนั้นในกรณีที่ออบเจ็คมีค่าความสามารถในการครอบคลุมสูงสุดมากกว่า 1 ออบเจ็ค จะทำการสุ่มเลือกแทน และเรียกวิธีนี้ว่า "วิธี MCS แบบสุ่ม"

ส่วนการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จากวิธี MCS (แบบดั้งเดิม) GAs และ HGAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1 ชุดข้อมูลที่ใช้ในการทดลอง

ชุดข้อมูลที่ใช้ในวิทยานิพนธ์ฉบับนี้นำมาจากฐานข้อมูลมาตรฐานของ "UCI Machine Learning Repository" [18] ซึ่งเป็นชุดข้อมูลที่เก็บมาจากข้อมูลจริง ซึ่งมีค่าคุณลักษณะเฉพาะที่เป็นจำนวนจริง และมีค่าคุณลักษณะครบทุกถ้วนข้อมูลเท่านั้น โดยถือว่าเป็นชุดข้อมูลที่มีความถูกต้องและไม่มีข้อมูลที่ผิดพลาด (Noise)

โดยแบ่งชุดข้อมูลเป็น 3 ชุด ตามจำนวนข้อมูลของชุดข้อมูล ซึ่งได้แก่ ชุดข้อมูลขนาดเล็ก ชุดข้อมูลขนาดกลาง และชุดข้อมูลขนาดใหญ่ โดยแต่ละชุดข้อมูลประกอบไปด้วย 2 ชุดข้อมูลย่อย ดังนี้

5.1.1 ชุดข้อมูลขนาดเล็ก

5.1.1.1 ชุดข้อมูล IRIS

ชุดข้อมูล IRIS เป็นชุดข้อมูลของดอกไอริช ใช้ในการจำแนกชนิดของดอกไอริช โดยพิจารณาจากความกว้าง และความยาวของกลีบเลี้ยง และกลีบดอก มีจำนวนข้อมูลทั้งหมด 150 ตัว มีค่าคุณลักษณะ 4 ค่า และแบ่งเป็น 3 ชนิด คือ Setosa, Versicolour และ Virginica

5.1.1.2 ชุดข้อมูล ECOLI

ชุดข้อมูล ECOLI เป็นชุดข้อมูลของแบคทีเรียอีโคไล ใช้ในการจำแนกชนิดของแบคทีเรียอีโคไล โดยพิจารณาจากรายละเอียดต่างๆ มีจำนวนข้อมูลทั้งหมด 336 ตัว มีค่าคุณลักษณะ 7 ค่า และแบ่งเป็น 8 ชนิด

ตารางที่ 5.1 รายละเอียดชุดข้อมูลขนาดเล็ก

ชุดข้อมูล	จำนวนข้อมูล	คุณลักษณะเฉพาะ	ชนิด
IRIS	150	4	3
ECOLI	336	7	8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.2 ชุดข้อมูลขนาดกลาง

5.1.2.1 ชุดข้อมูล CREDIT

ชุดข้อมูล CREDIT เป็นชุดข้อมูลของสินเชื่อ ใช้ในการตัดสินใจว่าควรจะอนุมัติสินเชื่อให้ลูกค้าหรือไม่ โดยพิจารณาจากรายละเอียดต่างๆ มีจำนวนข้อมูลทั้งหมด 1000 ตัว มีค่าคุณลักษณะ 84 ค่า และแบ่งเป็น 2 ชนิด คือ อนุมัติ และไม่อนุมัติ

5.1.2.2 ชุดข้อมูล YEAST

ชุดข้อมูล YEAST เป็นชุดข้อมูลของยีส ใช้ในการจำแนกชนิดของยีส โดยพิจารณาจากรายละเอียดต่างๆ มีจำนวนข้อมูลทั้งหมด 1484 ตัว มีค่าคุณลักษณะ 8 ค่า และแบ่งเป็น 10 ชนิด

ตารางที่ 5.2 รายละเอียดชุดข้อมูลขนาดกลาง

ชุดข้อมูล	จำนวนข้อมูล	คุณลักษณะเฉพาะ	ชนิด
CREDIT	1000	24	2
YEAST	1484	8	10

5.1.3 ชุดข้อมูลขนาดใหญ่

5.1.3.1 ชุดข้อมูล MUSK

ชุดข้อมูล MUSK เป็นชุดข้อมูลของน้ำหอม ใช้ในการตัดสินใจว่าน้ำหอมใหม่จะหอมถูกใจลูกค้าหรือไม่ โดยพิจารณาจากองค์ประกอบของน้ำหอม มีจำนวนข้อมูลทั้งหมด 6598 ตัว มีค่าคุณลักษณะ 166 ค่า และแบ่งเป็น 2 ชนิด คือ หอม และไม่หอม

5.1.3.2 ชุดข้อมูล PEN

ชุดข้อมูล PEN เป็นชุดข้อมูลของลายมือเขียน ใช้ในการจำแนกตัวเลขลายมือเขียน โดยพิจารณาจากรายละเอียดต่างๆ มีจำนวนข้อมูลทั้งหมด 6598 ตัว มีค่าคุณลักษณะ 166 ค่า และแบ่งเป็น 10 ชนิด คือ ตัวเลข 0 ถึง 9

ตารางที่ 5.3 รายละเอียดชุดข้อมูลขนาดใหญ่

ชุดข้อมูล	จำนวนข้อมูล	คุณลักษณะเฉพาะ	ชนิด
MUSK	6598	166	2
PEN	10092	16	10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ผลการทดลอง

5.2.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ที่ได้จากวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม เพื่อวิเคราะห์ว่าลำดับการเลือกที่แตกต่างกันจะส่งผลอย่างไรกับโปรโตไทป์ผลลัพธ์

วิธี MCS แบบดั้งเดิมมีขั้นตอนการทำงานอ้างอิงตามหัวข้อที่ 3.1 ในบทที่ 3 กล่าวคือในกรณีที่มีออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดมากกว่า 1 ออบเจ็ค วิธี MCS แบบดั้งเดิมจะทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดตัวแรก (เมื่อพิจารณาตามลำดับข้อมูลในชุดข้อมูล) เป็นโปรโตไทป์เสมอ

ส่วนวิธี MCS แบบสุ่มมีขั้นตอนการทำงานอ้างอิงตามหัวข้อที่ 3.1 ในบทที่ 3 เช่นกัน แต่ในกรณีที่มีออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดมากกว่า 1 ออบเจ็ค วิธี MCS แบบสุ่มจะทำการสุ่มเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดจากออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดทั้งหมดด้วยโอกาสเท่าๆ กัน และเนื่องจากวิธี MCS แบบสุ่มเป็นการทำงานเชิงสุ่มที่ผลลัพธ์ในแต่ละครั้งอาจจะไม่เหมือนกัน จึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

ผลการทดลองเพื่อเปรียบเทียบจำนวนผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม แสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ และเวลาการทำงาน (เวลาการทำงานจะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที) ของทั้ง 6 ชุดข้อมูล โดยมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล IRIS

ตารางที่ 5.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล IRIS

MCS แบบดั้งเดิม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	15	00:00:00.03

(ก)

MCS แบบสุ่ม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	15	00:00:00.03
2	15	00:00:00.05
3	14	00:00:00.03
4	16	00:00:00.03
5	15	00:00:00.05
ค่าเฉลี่ย	15.0	00:00:00.04

(ข)

จากตารางที่ 5.4 (ก) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบดั้งเดิมมีจำนวนโปรโตไทป์ผลลัพธ์ 15 ตัว (คิดเป็นร้อยละ 10.0 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงาน 0.03 วินาที ส่วนในตาราง 5.4 (ข) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบสุ่มมีจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ย 15.0 ตัว (คิดเป็นร้อยละ 10.0 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงานเฉลี่ย 0.04 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล ECOLI

ตารางที่ 5.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล ECOLI

MCS แบบดั้งเดิม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	101	00:00:00.47

(ก)

MCS แบบสุ่ม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	102	00:00:00.48
2	100	00:00:00.58
3	100	00:00:00.59
4	100	00:00:00.58
5	102	00:00:00.47
ค่าเฉลี่ย	100.8	00:00:00.54

(ข)

จากตารางที่ 5.5 (ก) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบดั้งเดิมมีจำนวนโปรโตไทป์ผลลัพธ์ 101 ตัว (คิดเป็นร้อยละ 30.05 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงาน 0.47 วินาที ส่วนในตารางที่ 5.5 (ข) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบสุ่มมีจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ย 100.8 ตัว (คิดเป็นร้อยละ 29.94 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงานเฉลี่ย 0.54 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล CREDIT

ตารางที่ 5.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล CREDIT

MCS แบบดั้งเดิม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	474	00:00:37.75

(ก)

MCS แบบสุ่ม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	471	00:00:31.64
2	479	00:00:31.72
3	473	00:00:32.23
4	471	00:00:32.19
5	476	00:00:33.17
ค่าเฉลี่ย	474.0	00:00:32.19

(ข)

จากตารางที่ 5.6 (ก) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบดั้งเดิมมีจำนวนโปรโตไทป์ผลลัพธ์ 474 ตัว (คิดเป็นร้อยละ 47.4 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงาน 37.75 วินาที ส่วนในตารางที่ 5.6 (ข) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบสุ่มมีจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ย 474.0 ตัว (คิดเป็นร้อยละ 47.4 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงานเฉลี่ย 32.19 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล YEAST

ตารางที่ 5.7 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล YEAST

MCS แบบดั้งเดิม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	886	00:02:43.41

(ก)

MCS แบบสุ่ม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	880	00:03:57.03
2	880	00:02:44.17
3	879	00:03:20.66
4	880	00:02:43.14
5	879	00:03:20.78
ค่าเฉลี่ย	879.6	00:03:13.16

(ข)

จากตารางที่ 5.7 (ก) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบดั้งเดิมมีจำนวนโปรโตไทป์ผลลัพธ์ 886 ตัว (คิดเป็นร้อยละ 59.70 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงาน 2 นาที 43.41 วินาที ส่วนในตารางที่ 5.7 (ข) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบสุ่มมีได้จำนวนโปรโตไทป์ผลลัพธ์เฉลี่ย 879.6 ตัว (คิดเป็นร้อยละ 59.27 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงานเฉลี่ย 3 นาที 13.16 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล MUSK

ตารางที่ 5.8 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล MUSK

MCS แบบดั้งเดิม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	828	00:21:51.00

(ก)

MCS แบบสุ่ม		
การทดลองครั้งที่	จำนวนโปรโตไทป์	เวลา
1	820	00:24:26.08
2	833	00:25:02.28
3	834	00:19:22.89
4	824	00:16:54.27
5	821	00:21:39.70
ค่าเฉลี่ย	826.4	00:21:29.04

(ข)

จากตารางที่ 5.8 (ก) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบดั้งเดิมมีจำนวนโปรโตไทป์ผลลัพธ์ 828 ตัว (คิดเป็นร้อยละ 12.55 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงาน 21 นาที 51.00 วินาที ส่วนในตารางที่ 5.8 (ข) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบสุ่มมีจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ย 826.4 ตัว (คิดเป็นร้อยละ 12.52 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงานเฉลี่ย 21 นาที 29.04 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล PEN

ตารางที่ 5.9 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่มกับชุดข้อมูล PEN

MCS แบบดั้งเดิม		
การทดลองครั้งที่	**จำนวนโปรโตไทป์**	**เวลา**
1	322	00:17:07.61

(ก)

MCS แบบสุ่ม		
การทดลองครั้งที่	**จำนวนโปรโตไทป์**	**เวลา**
1	324	00:18:40.84
2	325	00:16:17.55
3	333	00:16:26.42
4	329	00:17:43.38
5	329	00:16:16.58
ค่าเฉลี่ย	328.0	00:13:32.28

(ข)

จากตารางที่ 5.9 (ก) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบดั้งเดิมมีจำนวนโปรโตไทป์ผลลัพธ์ 322 ตัว (คิดเป็นร้อยละ 3.19 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงาน 17 นาที 7.61 วินาที ส่วนในตารางที่ 5.9 (ข) จะเห็นว่าผลลัพธ์ของวิธี MCS แบบสุ่มมีจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ย 328.0 ตัว (คิดเป็นร้อยละ 3.25 เมื่อเทียบกับจำนวนข้อมูลทั้งหมด) โดยใช้เวลาทำงานเฉลี่ย 13 นาที 32.28 วินาที

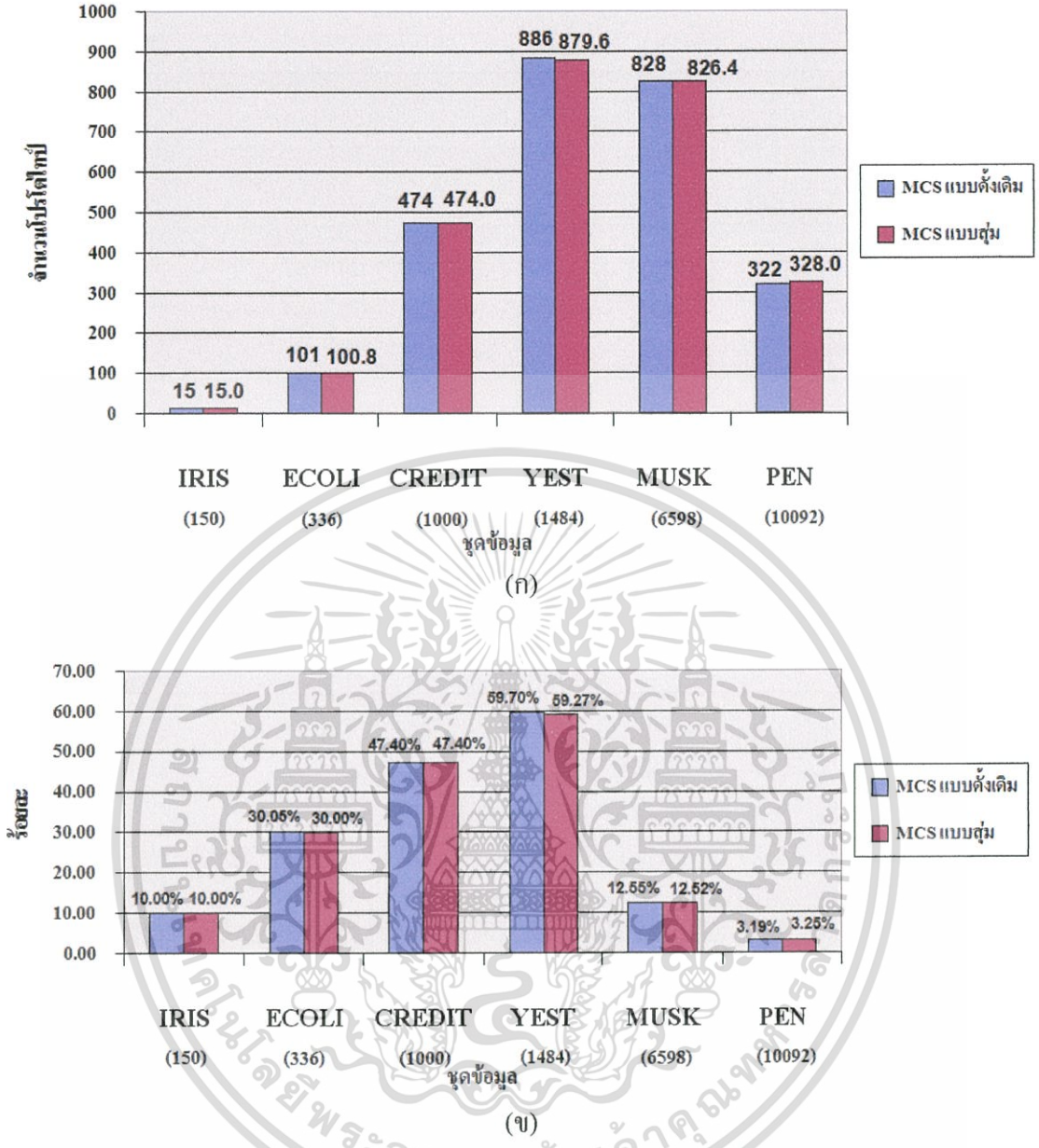
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.10 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม

		IRIS (150)	ECOLI (336)	CREDIT (1000)	YEAST (1484)	MUSK (6598)	PEN (10092)
MCS แบบดั้งเดิม	จำนวนโปรโตไทป์	15	101	474	886	828	322
	ร้อยละ	10.0	30.05	47.4	59.70	12.55	3.19
MCS แบบสุ่ม	จำนวนโปรโตไทป์ (เฉลี่ย)	15.0	100.8	474.0	879.6	826.4	328.0
	ร้อยละ	10.0	30.0	47.4	59.27	12.52	3.25

ตารางที่ 5.10 เป็นตารางแสดงสรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม โดยช่องจำนวนโปรโตไทป์แสดงจำนวนโปรโตไทป์ผลลัพธ์ของวิธี MCS แบบดั้งเดิม และวิธี MCS แบบสุ่ม โดยในส่วนช่องจำนวนออบเจ็คของโปรโตไทป์ผลลัพธ์ของวิธี MCS แบบสุ่มจะพิจารณาจากค่าเฉลี่ยของจำนวนออบเจ็คของโปรไทป์ผลลัพธ์จากการทดลองทั้ง 5 ครั้ง ของทั้ง 6 ชุดข้อมูล ส่วนช่องร้อยละแสดงร้อยละของจำนวนออบเจ็คของโปรโตไทป์ผลลัพธ์เทียบกับจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ และตัวเลขในวงเล็บใช้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม

จากรูปที่ 5.1 แสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม โดยรูป (ก) แสดงผลการทดลองในรูปของจำนวนโปรโตไทป์ผลลัพธ์ที่ได้ ส่วนรูป (ข) แสดงผลการทดลองในรูปของร้อยละของจำนวนโปรโตไทป์ผลลัพธ์เมื่อเทียบกับจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ ซึ่งจะเห็นว่าทั้งวิธี MCS แบบดั้งเดิมและวิธี MCS แบบสุ่มให้ผลลัพธ์ที่ใกล้เคียงกันมาก หมายความว่าในกรณีที่มีออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดมากกว่า 1 ออบเจ็ค ไม่ว่าจะเลือกออบเจ็คที่มีความสามารถในการครอบคลุมสูงสุดออบเจ็คไหนเป็นโปรโตไทป์ก็จะได้ผลลัพธ์ไม่แตกต่างกัน หรือกล่าวคือวิธี MCS แบบดั้งเดิมกับวิธี MCS แบบสุ่ม “ไม่มีความแตกต่างอย่างมีนัยสำคัญ”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ของวิธี MCS (แบบดั้งเดิม) GAs และ HGAs

วิธี MCS ที่ใช้เป็นวิธี MCS แบบดั้งเดิมมีขั้นตอนการทำงานอ้างอิงตามหัวข้อที่ 3.1 ในบทที่ 3 กล่าวคือในกรณีที่มีออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดมากกว่า 1 ออบเจ็ค วิธี MCS แบบดั้งเดิมจะทำการเลือกออบเจ็คที่มีค่าความสามารถในการครอบคลุมสูงสุดตัวแรก (เมื่อพิจารณาตามลำดับข้อมูลในชุดข้อมูลนั้นๆ) เป็นโปรโตไทป์เสมอ

ส่วน GAs มีขั้นตอนการทำงานอ้างอิงตามหัวข้อที่ 3.2 ในบทที่ 3 และมีการกำหนดค่าตัวแปรต่างๆ ดังนี้

- จำนวนประชากร (N_{pop}) กำหนดให้มีค่าเท่ากับ “40”
- ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) กำหนดให้มีค่าเท่ากับ “0.8”
- ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) กำหนดให้มีค่าเท่ากับ “0.5”
- ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) กำหนดให้มีค่าเท่ากับ “0.015”

การกำหนดค่าตัวแปรต่างๆ ของ GAs ข้างต้นอาจจะไม่ใช้วิธีการกำหนดที่เหมาะสมกับทุกชุดข้อมูล แต่เมื่อพิจารณาจากผลการทดลองในภาคผนวก ฌ. ซึ่งสรุปได้ว่าไม่มีวิธีการกำหนดค่าตัวแปรต่างๆ ที่เหมาะสมกับทุกชุดข้อมูลได้ ดังนั้นวิทยานิพนธ์ฉบับนี้จึงนำเสนอ GAs แบบที่ใช้กำหนดค่าตัวแปรต่างๆ ข้างต้น

ส่วน HGAs มีขั้นตอนการทำงานอ้างอิงจากหัวข้อที่ 4.2 ในบทที่ 4 และมีการกำหนดค่าตัวแปรต่างๆ ดังนี้

- จำนวนประชากร (N_{pop}) กำหนดให้มีค่าเท่ากับ “10”
- ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) กำหนดให้มีค่าเท่ากับ “0.8”
- ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) กำหนดให้มีค่าเท่ากับ “0.5”
- ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) กำหนดให้มีค่าเท่ากับ “0.015”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากการเลือกโปรโตไทป์เพื่อลดจำนวนโปรโตไทป์นั้นจะทำการลดจำนวนโปรโตไทป์ลงให้มากที่สุดเท่าที่จะหาได้ ดังนั้นจึงไม่สามารถทำการกำหนดเกณฑ์เงื่อนไขการสิ้นสุดการทำงานให้กับ GAs และ HGAs ได้ และเนื่องจากการทำงานใน 1 เจเนอเรชั่นของ HGAs นั้นมีความซับซ้อนกว่า GAs มาก จึงไม่เหมาะสมที่จะทำการกำหนดให้ GAs และ HGAs ทำงานในจำนวนเจเนอเรชั่นที่เท่ากัน ดังนั้นจึงทำการกำหนดให้ในแต่ละการทดลองของ GAs และ HGAs จะทำงานเท่ากันเป็นเวลา 12 ชั่วโมง

โดยการเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ของ GAs และ HGAs จะทำการเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์กันใน 3 กรณี ดังนี้

1. กรณีที่สอดคล้อง 100% คือ กรณีที่การจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิง สามารถจำแนกข้อมูลในชุดข้อมูลฝึกสอน ได้ถูกต้องทุกตัว

2. กรณีที่สอดคล้อง 99% คือ กรณีที่การจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิง สามารถจำแนกข้อมูลในชุดข้อมูลฝึกสอน ได้ถูกต้องอย่างน้อย 99% เมื่อเทียบกับจำนวนข้อมูลทั้งหมด

3. กรณีที่สอดคล้อง 98% คือ กรณีที่การจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิง สามารถจำแนกข้อมูลในชุดข้อมูลฝึกสอน ได้ถูกต้องอย่างน้อย 98% เมื่อเทียบกับจำนวนข้อมูลทั้งหมด

ผลการทดลองของวิธี MCS แสดงในตางรางที่ 5.11 โดยช่องจำนวนโปรโตไทป์แสดงจำนวนโปรโตไทป์ผลลัพธ์ของวิธี MCS (แบบดั้งเดิม) ของทั้ง 6 ชุดข้อมูล ส่วนช่องร้อยละแสดงร้อยละของจำนวนโปรโตไทป์ผลลัพธ์เทียบกับจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ และตัวเลขในวงเล็บใช้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

ส่วนผลการทดลองของ GAs และ HGAs แสดงในตางรางที่ 5.12 และ 5.13 ตามลำดับ โดยจำนวนโปรโตไทป์ผลลัพธ์ที่แสดงเป็นค่าเฉลี่ย (เนื่องจาก GAs และ HGAs เป็นวิธีการเชิงสุ่มซึ่งผลลัพธ์ที่ได้ในแต่ละการทำงานจะให้ผลลัพธ์ไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย) ของจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ (เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที) ของทั้ง 6 ชุดข้อมูล เมื่อทำงานเป็นเวลา 12 ชั่วโมง และตัวเลขในวงเล็บใช้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

รายละเอียดผลการทดลองของ GAs และ HGAs อยู่ในภาคผนวก ก. และภาคผนวก ข. ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.11 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของวิธี MCS

ชุดข้อมูล	จำนวนโปรโตไทป์	เวลา
IRIS (150)	15	00:00:00.31
ECOLI (336)	101	00:00:00.47
CREDIT (1000)	474	00:00:37.75
YEAST (1484)	886	00:02:43.41
MUSK (6598)	828	00:21:51.00
PEN (10092)	322	00:17:07.61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.12 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	- เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	11.2	01:18:00.01	9.4	00:59:40.03	7.2	01:20:08.34
ECOLI (336)	89.2	02:06:36.59	84.4	02:35:08.89	82.8	02:44:58.77
CREDIT (1000)	461.6	10:37:49.98	453.2	10:38:10.46	445.0	10:07:33.34
YEAST (1484)	1186.2	07:22:28.55	1068.2	10:18:18.24	1059.2	10:59:40.44
MUSK (6598)	3916.0	08:38:21.67	3175.8	11:02:21.26	3175.8	11:02:21.26
PEN (10092)	3318.4	11:34:09.38	3291.4	11:23:44.75	3291.4	11:23:44.75

จากตารางที่ 5.12 จะเห็นว่าในชุดข้อมูล IRIS, ชุดข้อมูล ECOLI และชุดข้อมูล CREDIT นั้น GAs ให้ผลลัพธ์ที่ดีกว่าวิธี MCS แต่ในชุดข้อมูล YEAST, ชุดข้อมูล MUSK และชุดข้อมูล PEN นั้นวิธี MCS ให้ผลลัพธ์ที่ดีกว่า GAs เป็นเพราะในชุดข้อมูลที่มีข้อมูลจำนวนมาก จะมีปริภูมิคำตอบขนาดมหาศาล ทำให้ GAs ไม่สามารถเฟ้นสุ่มหาคำตอบที่ดีได้ (ภายในเวลาที่กำหนด) จึงสรุปได้ว่ายิ่งชุดข้อมูลมีข้อมูลจำนวนมากขึ้น GAs ก็จะให้ผลลัพธ์ที่แย่ลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.13 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:04.18	9.0	00:00:11.23	7.0	00:14:14.76
ECOLI (336)	87.8	01:39:09.46	82.8	03:37:37.01	80.4	03:20:53.29
CREDIT (1000)	457.6	00:01:32.69	451.2	03:48:37.40	435.4	01:40:17.66
YEAST (1484)	869.0	00:07:07.28	869.0	00:07:07.28	853.2	00:05:02.13
MUSK (6598)	792.0	02:33:41.32	761.0	02:29:33.59	749.6	01:58:10.14
PEN (10092)	311.0	01:46:56.29	282.4	01:36:15.39	282.4	01:36:15.39

จากตารางที่ 5.13 จะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่า GAs ในทุกชุดข้อมูล และทุกกรณีความสอดคล้อง โดยเฉพาะอย่างยิ่งในชุดข้อมูลที่มีข้อมูลจำนวนมาก เช่น ชุดข้อมูล MUSK และชุดข้อมูล PEN อีกทั้งเวลาที่ได้ผลลัพธ์ก็ยังใช้เวลาน้อยกว่า (เมื่อเทียบกับ GAs) และเมื่อเทียบกับวิธี MCS ก็จะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่าวิธี MCS ทุกชุดข้อมูล โดยเฉพาะอย่างยิ่งในชุดข้อมูล IRIS, ชุดข้อมูล ECOLI และชุดข้อมูล CREDIT นั้น HGAs ให้ผลลัพธ์ที่ดีกว่าวิธี MCS อย่างชัดเจน ส่วนในชุดข้อมูล YEAST, ชุดข้อมูล MUSK และชุดข้อมูล PEN นั้นแม้ว่า HGAs จะให้ผลลัพธ์ที่ดีกว่าไม่มากนัก แต่เมื่อพิจารณารายละเอียดการทดลองทั้ง 5 ครั้งของ HGAs ก็จะเห็นว่าให้ผลลัพธ์ดีกว่าวิธี MCS ทั้ง 5 ครั้ง

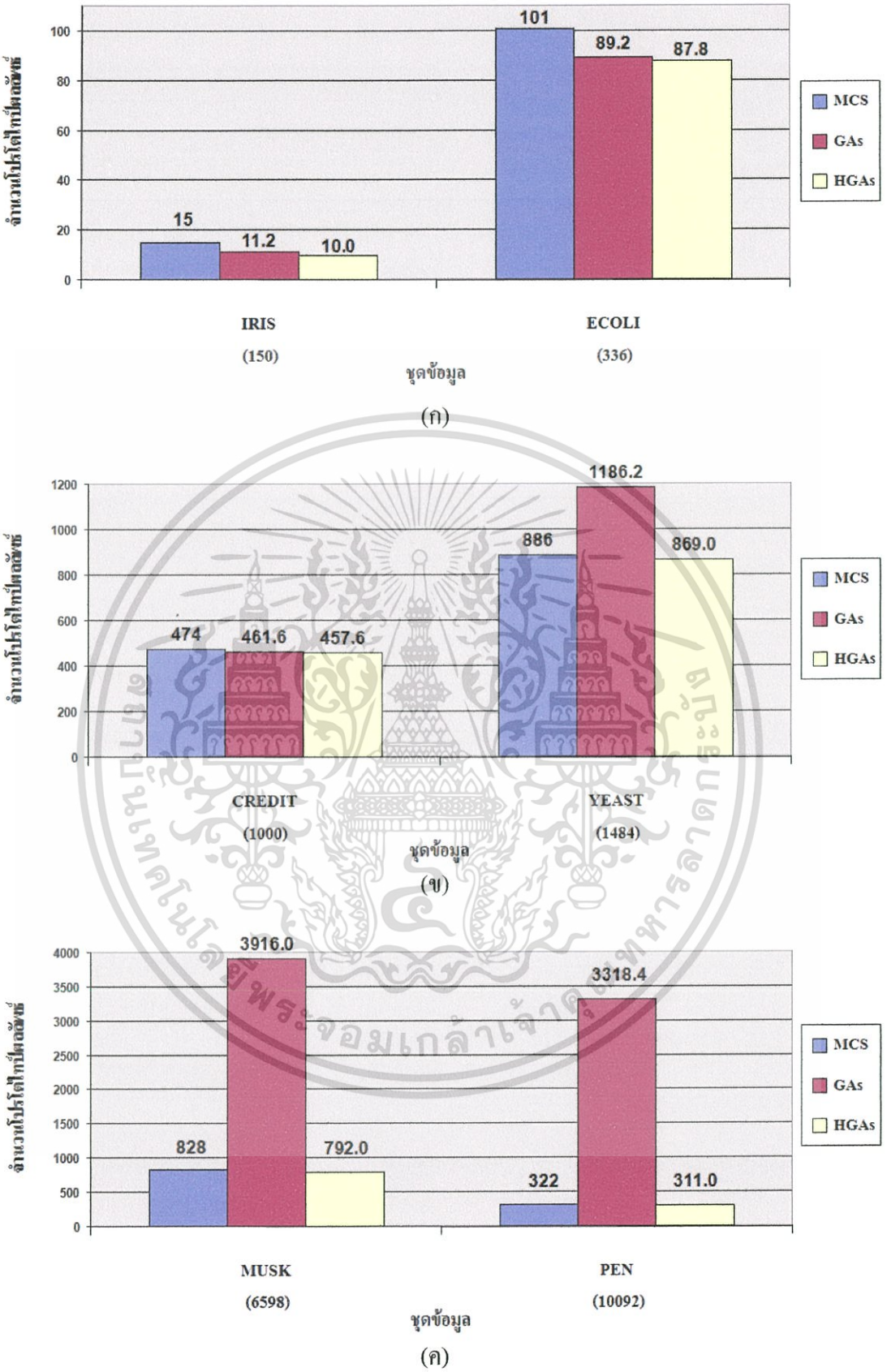
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.14 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ

ชุดข้อมูล	MCS (แบบดั้งเดิม)		GAs (100%)		HGAs (100%)	
	จำนวนโปรโตไทป์ (เฉลี่ย)	ร้อยละ	จำนวนโปรโตไทป์ (เฉลี่ย)	ร้อยละ	จำนวนโปรโตไทป์ (เฉลี่ย)	ร้อยละ
IRIS (150)	15	10.0	11.2	7.47	10.0	6.67
ECOLI (336)	101	30.06	89.2	26.55	87.8	26.13
CREDIT (1000)	474	47.4	461.6	46.16	457.6	45.76
YEAST (1484)	886	59.70	1186.2	79.93	869.0	58.56
MUSK (6598)	828	12.55	3916.0	59.35	792.0	12.0
PEN (10092)	322	3.19	3318.4	32.88	311.0	3.08

จากตารางที่ 5.14 เป็นตารางสรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ ของทุกชุดข้อมูล โดยช่องจำนวนโปรโตไทป์แสดงจำนวนโปรโตไทป์ผลลัพธ์ของวิธี MCS (แบบดั้งเดิม), GAs (เฉพาะกรณีสอดคล้อง 100%) และ HGAs (เฉพาะกรณีสอดคล้อง 100%) โดยในส่วนจำนวนโปรโตไทป์ผลลัพธ์ของ GAs และ HGAs จะพิจารณาจากค่าเฉลี่ยของจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุดที่ได้จากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 12 ชั่วโมง) ของทั้ง 6 ชุดข้อมูล ส่วนช่องร้อยละแสดงร้อยละของจำนวนโปรโตไทป์ผลลัพธ์เทียบกับจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ ส่วนตัวเลขในวงเล็บใช้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ (รายละเอียดผลการทดลองแต่ละครั้งแสดงในภาคผนวก ค.)

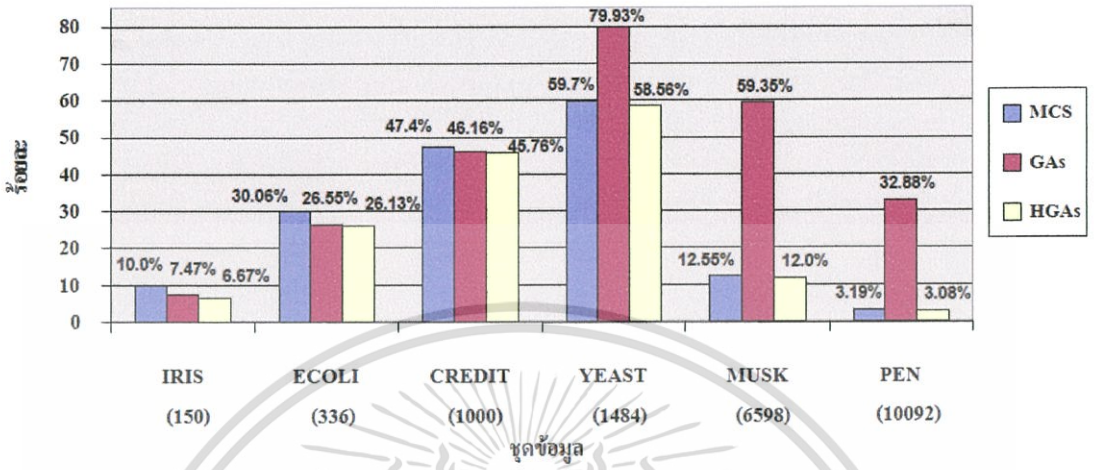
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ (เฉพาะกรณีสอดคล้อง 100%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 5.2 เป็นการแสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ โดยรูป (ก) แสดงผลการทดลองชุดข้อมูลขนาดเล็ก ส่วนรูป (ข) แสดงผลการทดลองชุดข้อมูลขนาดกลาง และรูป (ค) แสดงผลการทดลองชุดข้อมูลขนาดใหญ่



รูปที่ 5.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ ของทุกชุดข้อมูล (ในรูปแบบของร้อยละเมื่อเทียบกับจำนวนข้อมูลในแต่ละชุดข้อมูล)

รูปที่ 5.3 เป็นการแสดงสรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ ของทุกชุดข้อมูล (ในรูปแบบของร้อยละเมื่อเทียบกับจำนวนข้อมูลในชุดข้อมูล) โดยแนวแกนตั้งจะแสดงถึงร้อยละของจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จากแต่ละวิธีเทียบกับจำนวนข้อมูลทั้งหมดในแต่ละชุดข้อมูล ส่วนแนวแกนนอนแสดงถึงชุดข้อมูลที่ใช้ในการทดลองและตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

เมื่อพิจารณาชุดข้อมูลที่มีข้อมูลจำนวนไม่มาก อย่างเช่น ชุดข้อมูล IRIS, ชุดข้อมูล ECOLI และชุดข้อมูล CREDIT จะเห็นว่า GAs ให้ผลลัพธ์ที่ดีกว่าวิธี MCS เพราะปริภูมิคำตอบยังมีขนาดไม่ใหญ่มากนัก GAs เพียงอย่างเดียวยังสามารถเฟ้นสุ่มหาคำตอบที่ดีได้ แต่เมื่อพิจารณาชุดข้อมูลที่มีจำนวนจ้อมูลมากขึ้น เช่น ชุดข้อมูล YEAST, ชุดข้อมูล MUSK และชุดข้อมูล PEN วิธี MCS ให้ผลลัพธ์ที่ดีกว่า GAs เพราะปริภูมิคำตอบมีขนาดมหาศาล GAs อย่างเดียวไม่สามารถสุ่มหาคำตอบที่ดีได้ แม้จะใช้เวลาทำงานนานถึง 12 ชั่วโมงก็ตาม ในทางกลับกัน HGAs ให้ผลลัพธ์ที่ดีกว่าทั้งวิธี MCS และ GAs ทั้งในชุดข้อมูลขนาดเล็ก ชุดข้อมูลขนาดกลาง และชุดข้อมูลขนาดใหญ่ และเมื่อเปรียบเทียบกับ GAs ก็จะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่า GAs ในกรณีสอดคล้อง 100%, กรณีสอดคล้อง 99% และกรณีสอดคล้อง 98% (เมื่อทำงานเป็นเวลา 12 ชั่วโมง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการทดลองของวิธี MCS ในตารางที่ 5.14 เป็นเครื่องยืนยันได้อย่างดีว่าผลลัพธ์ของวิธี MCS ไม่ใช่ผลลัพธ์ที่ดีที่สุด เพราะ HGAs ให้ผลลัพธ์ที่ดีกว่าในทุกชุดข้อมูล แต่ผลลัพธ์ของ HGAs ก็อาจจะไม่ใช่ผลลัพธ์ที่ดีที่สุดเช่นกัน แต่เมื่ออ้างอิงจากบทความที่มีชื่อว่า "Minimal Consistent Subset Selection as Integer Nonlinear Programming Problem" [19] ที่ประยุกต์เอาหลักการทางคณิตศาสตร์เข้ามาแก้ปัญหาการเลือกโปรโตไทป์ ซึ่งทำให้ทราบว่าผลลัพธ์ที่ดีที่สุดของชุดข้อมูล IRIS คือ โปรโตไทป์ผลลัพธ์ที่มีจำนวนโปรโตไทป์เท่ากับ "10" ตัว (โดยโปรโตไทป์ผลลัพธ์จะต้องสอดคล้องกับชุดข้อมูลฝึกสอนด้วย) และเมื่อพิจารณารายละเอียดการทดลองของ GAs และ HGAs ในภาคผนวก ก. และภาคผนวก ข. (ตามลำดับ) ก็จะเห็นว่า HGAs สามารถหาผลลัพธ์ที่ดีที่สุดได้ทั้ง 5 ครั้งที่ทำการทดลอง ในขณะที่ GAs สามารถหาผลลัพธ์ที่ดีที่สุดได้เพียง 2 ครั้งเท่านั้น

ดังนั้นจึงแสดงตัวอย่างการทดลองของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ เพื่อแสดงให้เห็นถึงความสำคัญของกระบวนการต่างๆ ใน HGAs ที่ต้องทำงานร่วม จนทำให้ได้ผลลัพธ์ที่ดีที่สุด โดยจะแสดง 2 ตัวอย่าง ดังนี้

1. ตัวอย่างการทดลองที่ 1 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้
2. ตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้

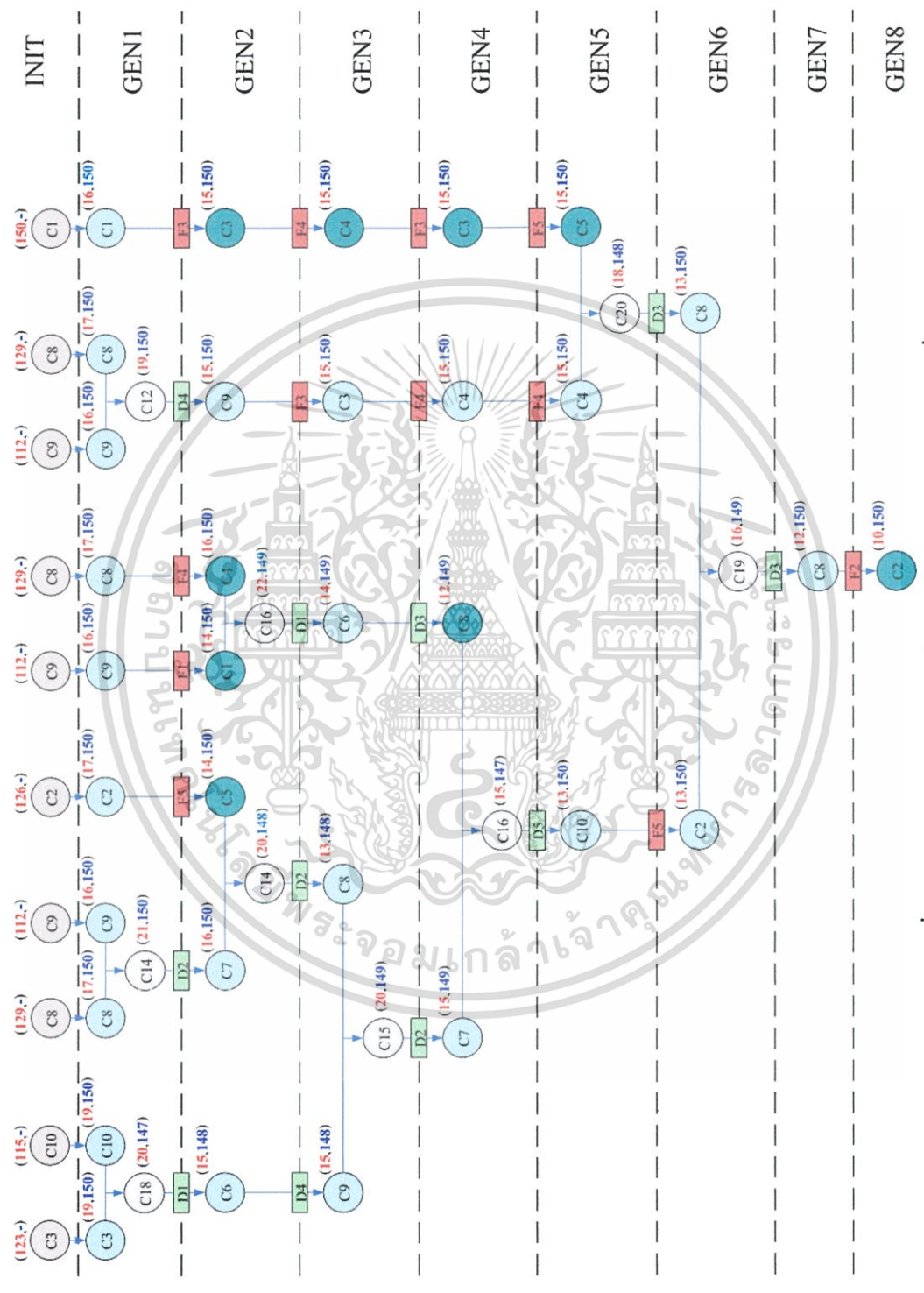
ตัวอย่างการทดลองที่ 1 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้จะแสดงในบทนี้ ส่วนตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้จะแสดงในภาคผนวก ง.

HGAs ที่ใช้ในการทดลองมีขั้นตอนการทำงานอ้างอิงจากหัวข้อที่ 4.2 ในบทที่ 4 และมีการกำหนดค่าตัวแปรต่างๆ ดังนี้

- จำนวนประชากร (N_{pop}) กำหนดให้มีค่าเท่ากับ "10"
- ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) กำหนดให้มีค่าเท่ากับ "0.8"
- ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) กำหนดให้มีค่าเท่ากับ "0.5"
- ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) กำหนดให้มีค่าเท่ากับ "0.015"

ตัวอย่างการทดลองที่ 1 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ เป็นผลการทดลองของ HGAs สามารถหาผลลัพธ์ที่มีจำนวนโปรโตไทป์เท่ากับ "10" ได้ในเจเนอเรชั่นที่ "8" โดยแสดงในลักษณะแผนภูมิแสดงลำดับเครือญาติ (Family Tree) เพื่อวิเคราะห์ที่มาของผลลัพธ์ที่ดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

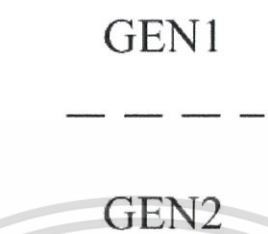


รูปที่ 5.4 แผนภูมิแสดงลำดับเครื่องญาติของตัวอย่างการทดลองที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

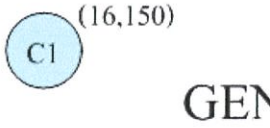
โดยกำหนดให้

- เส้นประ แสดงถึงขอบเขตระหว่างแต่ละเจเนอเรชั่น โดยจากตัวอย่างการทดลองที่ 1 จะเริ่มต้นทำงานตั้งแต่กระบวนการกำหนดค่าเริ่มต้น ไปจนถึงเจเนอเรชั่นที่ 8 ซึ่งเป็นเจเนอเรชั่นที่หาผลลัพธ์ที่ดีที่สุดได้ ดังแสดงตัวอย่างในรูปที่ 5.5



รูปที่ 5.5 ขอบเขตระหว่างเจเนอเรชั่นที่ 1 กับเจเนอเรชั่นที่ 2

- รูปวงกลม "Cn" โดยตัวอักษร "C" แสดงถึงโครโมโซมในแต่ละเจเนอเรชั่น ส่วน "n" แสดงถึงลำดับของโครโมโซมในเจเนอเรชั่นนั้นๆ และตัวเลขในวงเล็บข้างรูปวงกลมแสดงถึงค่า *Size*(S) และค่า *Accuracy*(S) ของโครโมโซมนั้นๆ ตามลำดับ โดยจากตัวอย่างการทดลองที่ 1 ที่กำหนดให้มีประชากรจำนวน 10 ตัว ดังนั้นในแต่ละเจเนอเรชั่นก็จะประกอบไปด้วยโครโมโซมจำนวน 10 ตัว (C1 ถึง C10) และจะมีโครโมโซมลูกที่เกิดจากกระบวนการทางพันธุกรรมอีก 10 ตัว (C11 ถึง C20) โดยที่ในเจเนอเรชั่นที่ 1 ลำดับของโครโมโซมจะเกิดจากการสุ่มจากกระบวนการกำหนดค่าเริ่มต้น แต่ตั้งแต่เจเนอเรชั่นที่ 2 เป็นต้นไป โครโมโซมลำดับที่ 1 ถึงโครโมโซมลำดับที่ 5 จะเรียงลำดับตามค่าความเหมาะสมจากมากไปน้อย (เพราะมาจากกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความเหมาะสมของเจเนอเรชั่นก่อนหน้า) ส่วนโครโมโซมลำดับที่ 6 ถึงโครโมโซมลำดับที่ 10 จะเรียงลำดับตามค่าความหลากหลายจากมากไปน้อย (เพราะมาจากกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความหลากหลายของเจเนอเรชั่นก่อนหน้า) โดยลำดับของโครโมโซมจะถูกกำหนดใหม่ในทุกเจเนอเรชั่น ดังแสดงตัวอย่างในรูปที่ 5.6



รูปที่ 5.6 โครโมโซมลำดับที่ 1 ในเจเนอเรชั่นที่ 1 (มีค่า *Size*(S)=16 และค่า *Accuracy*(S)=150)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- รูปสี่เหลี่ยมผืนผ้าพื้นแดง "Fn" โดยตัวอักษร "F" แสดงถึงลักษณะการถูกเลือกโดยกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความเหมาะสมในการเลือก (Fitness Selection) ส่วน "n" แสดงถึงลำดับของการถูกเลือกของโครโมโซมนั้นๆ โดยจากการทดลองข้างต้นกำหนดให้ประชากรมีจำนวน 10 ตัว ดังนั้นในกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความเหมาะสมจะทำการเลือกโครโมโซมจำนวน 5 ตัวจากโครโมโซมทั้งหมดที่มีตามค่าความเหมาะสมเป็นประชากรในรุ่นถัดไป ดังนั้น "F1" จะเป็นการเลือกโครโมโซมที่มีค่าความเหมาะสมมากที่สุด และโครโมโซมที่เลือกได้ก็จะกลายเป็นโครโมโซมลำดับที่ 1 ในเจเนอเรชั่นถัดไป ส่วนโครโมโซมที่ได้จาก "F2" ก็จะกลายเป็นโครโมโซมลำดับที่ 2 และเป็นเช่นนี้ไปเรื่อยๆ จนถึงโครโมโซมลำดับที่ 5 ตัวอย่างแสดงดังรูปที่ 5.7

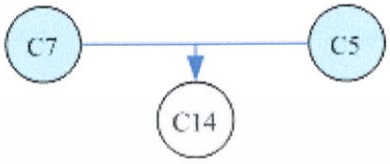
รูปที่ 5.7 กระบวนการเลือกประชากรในรุ่นถัดไปทำการเลือกโครโมโซมนี้ด้วยค่าความเหมาะสมเป็นลำดับที่ 3 เพื่อนำไปเป็นโครโมโซมลำดับที่ 3 ในเจเนอเรชั่นถัดไป

- รูปสี่เหลี่ยมผืนผ้าพื้นเขียว "Dn" โดยตัวอักษร "D" แสดงถึงลักษณะการถูกเลือกโดยกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความหลากหลายในการเลือก (Diversity Selection) ส่วน "n" แสดงถึงลำดับของการถูกเลือกของโครโมโซมนั้นๆ โดยจากการทดลองข้างต้นกำหนดให้ประชากรมีจำนวน 10 ตัว ดังนั้นในกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความหลากหลายจะทำการเลือกโครโมโซมอีกจำนวน 5 ตัวจากที่เหลือจากกระบวนการเลือกประชากรในรุ่นถัดไปที่ใช้ค่าความเหมาะสม ตามค่าความหลากหลายเป็นประชากรในรุ่นถัดไป ดังนั้น "D1" จะเป็นการเลือกโครโมโซมที่มีค่าความหลากหลายมากที่สุด และโครโมโซมที่เลือกได้ก็จะกลายเป็นโครโมโซมลำดับที่ 6 ในเจเนอเรชั่นถัดไป ส่วนโครโมโซมที่ได้จาก "D2" ก็จะกลายเป็นโครโมโซมลำดับที่ 7 และเป็นเช่นนี้ไปเรื่อยๆ จนถึงโครโมโซมลำดับที่ 10

รูปที่ 5.8 กระบวนการเลือกประชากรในรุ่นถัดไปทำการเลือกโครโมโซมนี้ด้วยค่าความหลากหลายเป็นลำดับที่ 4 เพื่อนำไปเป็นโครโมโซมลำดับที่ 9 ในเจเนอเรชั่นถัดไป

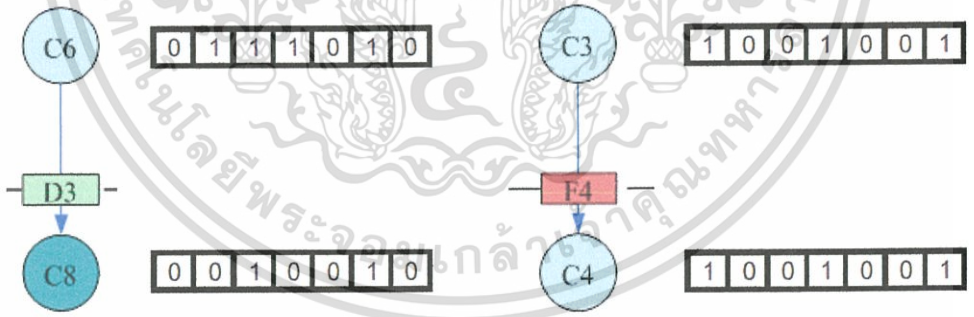
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- เส้นเชื่อมระหว่างคู่โครโมโซมในเจเนอเรชั่นเดียวกันแสดงถึง การเกิดกระบวนการทางพันธุกรรมที่มีคู่โครโมโซมนั้นเป็นโครโมโซมพ่อ-แม่ โดยจากการทดลองข้างต้นกำหนดให้ประชากรมีจำนวน 10 ตัว ดังนั้นจะเกิดกระบวนการทางพันธุกรรมขึ้น 5 ครั้ง ซึ่งแต่ละครั้งทำให้ได้โครโมโซมลูกจำนวน 2 ตัว ดังนั้นจะได้โครโมโซมลูกทั้งหมด 10 โครโมโซม ซึ่งก็จะกลายเป็นโครโมโซมลำดับที่ 11 ถึงโครโมโซมลำดับที่ 20 ตัวอย่างแสดงดังรูปที่ 5.9



รูปที่ 5.9 การเกิดกระบวนการทางพันธุกรรมระหว่างโครโมโซมลำดับที่ 7 กับโครโมโซลำดับที่ 5 เป็นผลให้ได้โครโมโซมลูกลำดับที่ 14

- สีของโครโมโซมตามลูกศรเมื่อผ่านจากเจเนอเรชั่นหนึ่งไปยังเจเนอเรชั่นหนึ่ง แสดงถึงการเปลี่ยนแปลงของโครโมโซมนั้นๆ โดยถ้าสีของโครโมโซมเปลี่ยนไปก็แสดงว่าโครโมโซมนั้นมีการเปลี่ยนแปลงเกิดขึ้น (โครงสร้างของโครโมโซมหลังผ่านกระบวนการ MCS มีความแตกต่างไปจากเดิม) แต่ถ้าสีของโครโมโซมไม่เปลี่ยนก็แสดงว่าโครโมโซมนั้นไม่มีการเปลี่ยนแปลง (โครงสร้างของโครโมโซมหลังผ่านกระบวนการ MCS เหมือนเดิม) ตัวอย่างแสดงดังรูปที่ 5.10



รูปที่ 5.10 ลักษณะการเปลี่ยนแปลงของโครโมโซม

จากรูปที่ 5.10 เป็นการแสดงถึงลักษณะการเปลี่ยนแปลงโครโมโซม โดยรูป (ก) เป็นลักษณะของโครโมโซมที่เปลี่ยนแปลง (สีของโครโมโซมเปลี่ยนไป) ซึ่งจะเห็นว่าค่าของบิตเปลี่ยนไป ส่วนรูป (ข) เป็นลักษณะของโครโมโซมที่ไม่เปลี่ยนแปลง (สีของโครโมโซมไม่เปลี่ยน) ซึ่งจะเห็นว่าค่าของบิตเหมือนเดิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 5.4 เป็นการแสดงโครงสร้างครอบครัวของตัวอย่างการทดลองที่ 1 ของ HGAs กับชุดข้อมูล IRIS โดยแสดงการพัฒนาของโครโมโซมตั้งกระบวนการกำหนดค่าเริ่มต้น โดยในแต่ละเจเนอเรชั่นจะมีโครโมโซมจำนวน 10 ตัว เรียงลำดับตั้งแต่โครโมโซมลำดับที่ 1 ถึงโครโมโซมลำดับที่ 10 (C1 ถึง C10)

จากนั้นเกิดกระบวนการทางพันธุกรรมขึ้นตัวอย่างเช่น โครโมโซมลำดับที่ 3 กับโครโมโซมลำดับที่ 10 ทำให้ได้โครโมโซมลูกลำดับที่ 18 ขึ้นมา โดยในแต่ละเจเนอเรชั่นจะทำกระบวนการทางพันธุกรรม 5 ครั้ง ซึ่งทำให้ได้โครโมโซมลูก 10 โครโมโซม เรียงลำดับตั้งแต่โครโมโซมลำดับที่ 11 ถึงโครโมโซมลำดับที่ 20 (C11 ถึง C20)

จากนั้นกระบวนการเลือกประชากรรุ่นถัดไปจะทำการเลือกโครโมโซมเพียง 10 โครโมโซมจากโครโมโซมทั้งหมด 20 โครโมโซม (โครโมโซมพ่อ-แม่จำนวน 10 ตัว และโครโมโซมลูกจำนวน 10 ตัว) ไปเป็นประชากรในรุ่นถัดไป โดยจะทำการเลือกโครโมโซมด้วยค่าความเหมาะสมก่อนจำนวน 5 ตัว จากนั้นจึงเลือกโครโมโซมด้วยค่าความหลากหลายอีกจำนวน 5 ตัว โดยโครโมโซมที่เลือกด้วยค่าความเหมาะสมจำนวน 5 ตัวจะกลายเป็นโครโมโซมลำดับที่ 1 ถึงโครโมโซมลำดับที่ 5 ในเจเนอเรชั่นถัดไป (C1 ถึง C5) ส่วนโครโมโซมที่เลือกด้วยค่าความเหมาะสมอีกจำนวน 5 ตัวจะกลายเป็นโครโมโซมลำดับที่ 6 ถึงโครโมโซมลำดับที่ 10 ในเจเนอเรชั่นถัดไป (C6 ถึง C10)

การทำงานเป็นเช่นนี้ตั้งแต่เจเนอเรชั่นที่ 1 ถึงเจเนอเรชั่นที่ 8 ซึ่งเป็นเจเนเรชั่นที่ HGAs หาผลลัพธ์ที่ดีที่สุดได้ โดยโครโมโซมที่เป็นผลลัพธ์ที่ดีที่สุดก็คือโครโมโซมลำดับที่ 2 ในเจเนอเรชั่นที่ 8 ซึ่งเมื่อทำการพิจารณาโครงสร้างครอบครัวของตัวอย่างการทดลองที่ 1 สามารถสรุปข้อสังเกตได้ 3 ข้อดังนี้

1. จะสังเกตเห็นว่าโครโมโซมบรรพบุรุษของโครโมโซมที่เป็นผลลัพธ์ที่ดีที่สุดเกิดมาจากกระบวนการทางพันธุกรรมถึง 9 ตัว นั่นหมายความว่ากระบวนการทางพันธุกรรมเป็นสิ่งสำคัญที่นำไปสู่ผลลัพธ์ที่ดีที่สุด

2. จะสังเกตเห็นว่าโครโมโซมบรรพบุรุษของโครโมโซมที่เป็นผลลัพธ์ที่ดีที่สุดถูกเลือกไปเป็นประชากรในรุ่นถัดไปโดย “Diversity Selection” ถึง 11 ครั้ง หมายความว่าการใช้เทคนิค “Diversity” เป็นสิ่งสำคัญที่นำไปสู่ผลลัพธ์ที่ดีที่สุด (โดยเฉพาะเทคนิค “Diversity Selection”)

3. จะสังเกตเห็นว่าโครโมโซมลำดับที่ 2 ในเจเนอเรชั่นที่ 8 ซึ่งเป็นผลลัพธ์ที่ดีที่สุดมาจากการพัฒนาจากโครโมโซมลำดับที่ 8 ในเจเนอเรชั่นที่ 7 โดยตรง นั่นหมายความว่าในการค้นหาคำตอบของ HGAs นั้น แม้ว่า GAs จะเป็นตัวช่วยกระจายค้นหาคำตอบในบริเวณกว้าง แต่คำตอบที่ได้ก็จะมาจากกระบวนการ MCS ที่เป็นตัวกรองคำตอบที่ได้นั้นอีกทีนั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผลการทดลองและข้อเสนอแนะ

6.1 สรุปผลการทดลอง

วิทยานิพนธ์ฉบับนี้เป็นอีกตัวอย่างหนึ่งของการทำงานร่วมกันระหว่างการค้นหาแบบกว้าง (Global Search) กับการค้นหาเฉพาะที่ (Local Search) ในการประยุกต์ใช้ในการเลือกโปรโตไทป์ เพื่อใช้แก้ปัญหาที่เกิดขึ้นในการจำแนกข้อมูลโดยใช้ NN Rule ในกรณีที่ชุดข้อมูลฝึกสอนมีข้อมูลจำนวนมาก โดยนำวิธี MCS มาทำงานร่วมกับ GAs อีกทั้งยังมีการนำเทคนิค "Singleton" และเทคนิค "Diversity" มาประยุกต์ใช้เพื่อเพิ่มประสิทธิภาพการทำงานด้วย

ตามที่ได้อธิบายไปตั้งแต่ต้นว่าขั้นตอนการทำงานของการจำแนกข้อมูลโดยใช้ NN Rule ต้องทำการคำนวณเปรียบเทียบข้อมูลที่ต้องการทราบชนิดกับข้อมูลทุกตัวในชุดข้อมูลฝึกสอน ดังนั้นในกรณีที่ชุดข้อมูลฝึกสอนมีข้อมูลจำนวนมาก ก็จะเสียเวลาในการคำนวณมากตามไปด้วย อีกทั้งจำเป็นต้องใช้หน่วยความจำเป็นจำนวนมากเพื่อใช้ในการเก็บข้อมูลเหล่านั้น ดังนั้นถ้าสามารถเลือกใช้เพียงข้อมูลบางตัวที่จำเป็น และสามารถทดแทนข้อมูลทั้งหมดได้ ก็จะช่วยลดปัญหาเหล่านี้ได้ ซึ่งสามารถทำได้โดยกระบวนการ "การเลือกโปรโตไทป์" ที่รับชุดข้อมูลฝึกสอนเป็นอินพุต และให้เอาพุตเป็นชุดข้อมูลฝึกสอนที่ลดจำนวนข้อมูลลงแล้ว (ซึ่งเรียกว่าโปรโตไทป์ผลลัพธ์)

สิ่งสำคัญในการเลือกโปรโตไทป์ คือ โปรโตไทป์ผลลัพธ์ต้องมีประสิทธิภาพการจำแนกข้อมูลใกล้เคียงกับชุดข้อมูลฝึกสอน ซึ่งสามารถแสดงได้ในรูปของขอบเขตการตัดสินใจ แต่การตรวจสอบโดยตรงนั้นทำได้ยาก ดังนั้นจึงนำคุณสมบัติความสอดคล้องเข้ามาช่วย โดยเพียงแค่ตรวจสอบว่าการจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิงสามารถจำแนกข้อมูลในชุดข้อมูลฝึกสอนได้ถูกต้องทุกตัวหรือไม่ ซึ่งถ้าสามารถจำแนกได้ถูกต้องทุกตัวก็หมายความว่าขอบเขตการตัดสินใจของโปรโตไทป์ผลลัพธ์มีความใกล้เคียงกับชุดข้อมูลฝึกสอน ทำให้การตรวจสอบประสิทธิภาพการจำแนกข้อมูลนั้นง่าย และสะดวกขึ้น

การเลือกโปรโตไทป์เพื่อลดจำนวนข้อมูลอ้างอิงลงให้น้อยที่สุดเท่าที่จะทำได้ โดยที่ยังคงรักษาประสิทธิภาพการจำแนกใกล้เคียงเดิม เป็นปัญหาที่เรียกว่า "Minimal Consistent Set Problem" กล่าวคือเป็นปัญหาที่ต้องการค้นหาเซตของข้อมูล (โปรโตไทป์ผลลัพธ์) ที่มำนวนน้อยที่สุดที่มีคุณสมบัติความสอดคล้อง (สามารถจำแนกชุดข้อมูลฝึกสอนได้ถูกต้องทั้งหมด) บนปริภูมิคำตอบ โดยการเลือกโปรโตไทป์มีมากมายหลายวิธี แต่วิธีที่วิทยานิพนธ์นี้สนใจ และนำมาเป็นพื้นฐานได้แก่ วิธี MCS และ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดย GAs เป็นวิธีหาผลลัพธ์เชิงวิทยาศาสตร์ที่นำมาประยุกต์ใช้ในการเลือกโปรโตไทป์ โดยทำการหาคำตอบโดยอาศัย “กระบวนการทางพันธุกรรม” ที่ทำงานในระดับ “ปริภูมิคำตอบ” เปรียบเสมือนการเฟ้นสุ่ม “ค้นหา” คำตอบบริเวณต่างๆ บนปริภูมิคำตอบ และทำการเก็บคำตอบที่หาได้ไว้ โดยถ้าบริเวณใดมีคำตอบที่ดีอยู่ก็จะทำการขยับไปหาคำตอบ ณ บริเวณใกล้เคียง ในทางกลับกันถ้าบริเวณใดไม่มีคำตอบที่ดีอยู่ก็จะทำการย้ายไปหาคำตอบ ณ บริเวณอื่นโดยจะทำการค้นหาตามเงื่อนไขที่กำหนดไว้ ส่วนผลลัพธ์ ก็คือ คำตอบที่ดีที่สุดในบรรดาคำตอบทั้งหมดที่เคยหาได้นั่นเอง ดังนั้นถ้าชุดข้อมูลมีข้อมูลจำนวนมาก GAs ก็ต้องเฟ้นสุ่มค้นหาคำตอบบนปริภูมิคำตอบมีขนาดมหาศาล ซึ่งเป็นไปได้ยากที่ GAs เพียงอย่างเดียวจะสามารถหาคำตอบที่ดีได้

ส่วนวิธี MCS เป็นการเลือกโปรโตไทป์โดยอาศัยหลัก “ความครอบคลุม” ที่ทำงานในระดับ “ปริภูมิคุณลักษณะเฉพาะ” เปรียบเสมือนการ “กรอง” ออบเจ็คที่ซ้ำซ้อนออกโดยอาศัยความรู้ที่ได้จากการคำนวณความครอบคลุม ทำให้เหลือแต่ออบเจ็คที่จำเป็นไว้ โดยทำการกรองซ้ำไปซ้ำมาจนกว่าจะได้ผลลัพธ์ที่ไม่สามารถทำการกรองให้เล็กต่อไปได้อีก ซึ่งตามที่ได้แสดงให้เห็นในบทที่ 3 ว่าการเริ่มต้นการทำงานโดยการพิจารณาออบเจ็คทั้งหมดในชุดข้อมูลไม่ใช่วิธีที่ทำให้ได้ผลลัพธ์ที่ดีที่สุด แต่การเริ่มการทำงานโดยการพิจารณาเพียงบางออบเจ็คกลับให้ผลลัพธ์ที่ดีกว่า แต่ก็ไม่มีวิธีที่จะสามารถทราบได้แน่นอนว่าควรเริ่มการทำงานโดยการพิจารณาออบเจ็คใดบ้าง เนื่องมากจากปัญหาแบบนี้เป็นปัญหาแบบ “ขัดแย้งกันเอง” กล่าวคือการที่จะทราบได้ว่าควรจะกำหนดออบเจ็คใดบ้างในการเริ่มการทำงานนั้น ก็ต้องทราบก่อนว่ามีออบเจ็คใดบ้างเป็นเป็นโปรโตไทป์ ซึ่งก็คือผลลัพธ์ของการเลือกโปรโตไทป์นั่นเอง อีกทั้งข้อมูลในแต่ละชุดข้อมูลก็มีลักษณะการวางตัวบนปริภูมิคุณลักษณะเฉพาะที่แตกต่างกันออกไป ทำให้ไม่มีวิธีที่แน่นอนที่สามารถบอกได้ว่าควรกำหนดออบเจ็คใดบ้างในการเริ่มต้นการทำงาน

วิทยานิพนธ์นี้จึงนำเสนอ “การเลือกโปรโตไทป์โดยใช้วิธี MCS และเจเนติกอัลกอริธึม” ซึ่งเป็นการนำวิธี MCS และ GAs มาทำงานร่วมกัน และเรียกวิธีใหม่นี้ว่า “HGAs” โดยเป็นการนำข้อดีของแต่ละวิธีมาแก้ข้อด้อยของอีกวิธี กล่าวคือจากข้อด้อยของวิธี MCS ที่จะเริ่มต้นการทำงานที่ตายตัวที่อาจจะส่งผลให้ผลลัพธ์ที่ได้ไม่ใช่ผลลัพธ์ที่ดีที่สุด จึงนำ GAs เข้ามาช่วยสุ่มพิจารณาออบเจ็คที่จะใช้เป็นอินพุตให้แทน ในทางกลับกันจากข้อด้อยของ GAs ที่ใช้ได้ไม่ดีกับข้อมูลที่มีจำนวนข้อมูลมาก เนื่องมาจากปริภูมิคำตอบมีขนาดมหาศาล จึงเป็นไปได้ยากที่จะเฟ้นสุ่มหาคำตอบที่ดีได้ จึงนำความรู้จากหลัก “ความครอบคลุม” ของวิธี MCS มาช่วยไกด์ให้กับ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

HGAs สามารถพิจารณาได้ง่ายๆ เปรียบเสมือน GAs ที่มีวิธี MCS เป็นกระบวนการตัดต่อพันธุกรรม ทำให้เกิดการวิวัฒนาการแบบก้าวกระโดด ส่งผลให้ประชากรพัฒนาไปในทิศทางที่ถูกต้อง เปรียบเสมือนการค้นหาคำตอบที่ใช้ GAs ในการสุ่มค้นหาคำตอบในวงกว้าง ก่อนที่จะใช้วิธี MCS ทำการกรองคำตอบอีกที จากนั้น GAs ก็จะทำการสุ่มค้นหาคำตอบอีก โดยทำการขยับไปหาคำตอบ ณ บริเวณใกล้เคียงของบริเวณที่มีคำตอบที่ดี ส่วนบริเวณใดมีคำตอบที่ไม่ดีก็จะไม่ทำการค้นหาคำตอบบริเวณนั้นต่อไป

กระบวนการค้นหาของ HGAs มี GAs เป็นกระบวนการที่มีทั้งการ “เพิ่ม/ลด” โปรโตไทป์ เปรียบเสมือนการค้นหาคำตอบแบบกว้างที่มีการเปลี่ยนแปลงบริเวณในการค้นหาคำตอบอยู่ตลอด ส่วนกระบวนการ MCS เป็นกระบวนการที่มีเฉพาะการ “ลด” โปรโตไทป์อย่างเดียว เปรียบเสมือนการกรองโปรโตไทป์ออก โดยอาศัยความรู้ที่ได้จากการคำนวณความครอบคลุม ทำให้เหลือแต่โปรโตไทป์ที่จำเป็นไว้

ดังนั้นจึงสรุปได้ว่ากระบวนการค้นหาของ HGAs นั้นเกิดจากการทำงานอย่างมีประสิทธิภาพระหว่างวิธี MCS กับ GAs แบบพึ่งพาอาศัยซึ่งกันและกัน กล่าวคือทั้ง 2 วิธีต้องทำงานร่วมกัน และไม่สามารถแบ่งแยกได้ว่าผลลัพธ์ที่ได้นั้นมาจากวิธีใดวิธีหนึ่งเป็นสำคัญ แต่ได้มาจากการทำงานร่วมกันอย่างมีประสิทธิภาพของทั้ง 2 วิธี

จากผลการทดลองในบทที่ 5 หัวข้อที่ 5.2 จะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่าทั้งวิธี MCS และ GAs ในทุกกรณีความสอดคล้อง และในทุกชุดข้อมูล โดยเฉพาะอย่างยิ่งในชุดข้อมูล IRIS นั้น HGAs ให้โปรโตไทป์ผลลัพธ์ที่มีจำนวนเท่ากับ “10” ซึ่งเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (Global Optimum) โดยอ้างอิงจากบทความที่มีชื่อว่า “Minimal Consistent Subset Selection as Integer Nonlinear Programming Problem” [19] ที่ประยุกต์เอาหลักการทางคณิตศาสตร์เข้ามาแก้ปัญหาการเลือกโปรโตไทป์ ซึ่งจะเห็นว่า HGAs ให้โปรโตไทป์ผลลัพธ์ที่มีจำนวนเท่ากับ “10” ทั้ง 5 ครั้งที่ทำการทดลอง ในขณะที่ GAs ให้โปรโตไทป์ผลลัพธ์ที่มีจำนวนเท่ากับ “10” ได้เพียง 2 ครั้งจาก 5 ครั้งที่ทำการทดลอง แต่เวลาที่ใช้ในการทำงานก็นานกว่า HGAs อย่างชัดเจน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.2 ข้อเสนอแนะ

การเลือกโปรโตไทป์ในวิทยานิพนธ์ฉบับนี้ยึดคุณสมบัติความสอดคล้องเป็นหลัก กล่าวคือ การจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิงจะต้องจำแนกข้อมูลในชุดข้อมูลฝึกสอนได้ถูกต้องทั้งหมด ซึ่งจะทำให้โปรโตไทป์ผลลัพธ์มีความจำเพาะในการจำแนกข้อมูลในชุดข้อมูลฝึกสอนมากจนเกินไป (Over Fitting) ทำให้เมื่อนำโปรโตไทป์ผลลัพธ์ไปเป็นข้อมูลอ้างอิงในการจำแนกข้อมูลโดยใช้ NN Rule ในชุดข้อมูลอื่นอาจจะได้ประสิทธิภาพการจำแนกที่ไม่ดีเท่าที่ควร

วิธีการแก้ไข คือ หาวิธีตรวจสอบประสิทธิภาพการจำแนกระหว่างโปรโตไทป์ผลลัพธ์กับชุดข้อมูลฝึกสอนใหม่ หรืออาจจะยอมผ่อนปรนให้มีการการจำแนกข้อมูลผิดพลาดได้บ้าง เพื่อให้โปรโตไทป์ผลลัพธ์มีความเหมาะสมกับข้อมูลทั่วไปมากขึ้น (Generalization) หรืออาจจะใช้ชุดข้อมูลอื่นหลายๆ ชุดเป็นตัวทดสอบ โดยถ้าการจำแนกข้อมูลโดยใช้ NN Rule ที่มีชุดข้อมูลฝึกสอนเป็นข้อมูลอ้างอิงสามารถจำแนกข้อมูลได้จำนวนใกล้เคียงกับการจำแนกข้อมูลโดยใช้ NN Rule ที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิง ก็จะถือว่าว่าชุดข้อมูลฝึกสอนกับโปรโตไทป์ผลลัพธ์มีประสิทธิภาพการจำแนกใกล้เคียงกัน

อีกส่วนหนึ่งที่น่าสนใจ คือ การเพิ่มประสิทธิภาพกรทำงานให้กับ HGAs โดยการนำเทคนิคอื่นๆ เข้ามาใช้ เนื่องจากที่ได้อธิบายไปข้างต้นแล้วว่าการนำเทคนิค “Singleton” และเทคนิค “Diversity” เข้ามาใช้ใน HGAs ช่วยทำให้ประสิทธิภาพการทำงานดีขึ้น ดังนั้นอาจจะมีเทคนิคอื่นๆ ที่สามารถนำมาใช้ใน HGAs แล้วทำให้ประสิทธิภาพเพิ่มขึ้นอีก เช่น การกำหนดตัวแปรต่างๆ ของ HGAs ซึ่งจะเห็นว่าเป็นการกำหนดที่ตายตัวตั้งแต่ต้นจนจบการทำงาน โดยที่ไม่มีการเปลี่ยนแปลงเลย ซึ่งในบางกรณีการกำหนดตัวแปรเช่นนี้อาจจะไม่เหมาะสม เช่น เมื่อประชากรมีความคล้ายคลึงกันมากๆ ประกอบกับปริภูมิคำตอบที่มีขนาดใหญ่มากๆ ซึ่งลำพังเทคนิค “Diversity” เพียงอย่างเดียวก็ไม่อาจจะช่วยทำให้ประสิทธิภาพการทำงานดีขึ้นได้ ก็ควรจะทำการเพิ่มค่าความน่าจะเป็นในการกลายพันธุ์ หรือเพิ่มจำนวนประชากรขึ้น เพื่อเพิ่มความหลากหลายให้กับประชากร และเมื่อประชากรมีความหลากหลายในระดับหนึ่งแล้วก็ค่อยลดค่าความน่าจะเป็นในการกลายพันธุ์ หรือลดจำนวนประชากรลง เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] P.N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", Addison-Wesley, 2006.

[2] R.O. Duda and P.E. Hart, "Pattern Classification and Scene Analysis", Wiley Interscience Publication,1973.

[3] T.M. Cover and P.E. Hart, "Nearest Neighbor Pattern Classification", IEEE Transactions on Information Theory, Vol. IT-13, No. 1, January 1967, pp. 21-27.

[4] B.V. Dasarathy, "Minimal Consistent Set (MCS) Identification for Optimal Nearest Neighbor Decision Systems Design", IEEE Transactions on Systems, Man and Cybernetics, Vol. 24, No. 3, 1994, pp. 511-517.

[5] L.I. Kuncheva and J.C. Bezdek, "Nearest Prototype Classification: Clustering, Genetic Algorithms, or Random Search?", IEEE Transactions on Systems, Man and Cybernetics, Vol. 28, No. 1, 1998, pp. 160-164.

[6] R.A. Mollineda, F.J. Ferri, and E. Vidal. "Merged-based prototype selection for nearest neighbor classification", Proceedings of the 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI2000), Orlando, USA, July 2000, pp. 640-645.

[7] J.S. Sanchez, F. Pla and F.J. Ferri, "Prototype Election for the Nearest Neighbor Rule through Proximity Graphs", Pattern Recognition Letters 18, 1997, pp. 507-513.

[8] J.C. Bezdek, T.R. Reichherzer, G.S. Lim and Y. Attikiouzel, "Multiple-prototype classifier design", IEEE Transaction. on Systems., Man and Cybernetics - part B : cybernetics,Vol. 28, Feb. 1998 ,pp. 67-79

[9] T. Kohonen, "The self-organizing map", Proc. IEEE, Vol 78, no. 9, pp.1464-1480, 1990.

[10] G P.E. Hart, "The Condensed Nearest Neighbor Rule", IEEE Transactions on Information Theory, Vol. 14, 1968, pp. 515-516.

[11] V. Cerverón and, A. Fuertes, "Parallel Random Search and Tabu Search for the Minimal Consistent Subset Selection Problem", Randomization and Approximation Techniques in Computer Science, Vol. 1518, 1998, pp. 248-259

[12] V. Cerverón and F.J. Ferri, "Another Move Toward the Minimum Consistent Subset: A Tabu Search Approach to the Condensed Nearest Neighbor Rule", IEEE Transactions on Systems, Man and Cybernetics - part B : cybernetics,Vol. 31, no. 3, june 2001.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[13] R. Klein, "Concrete and Abstract Voronoi Diagram", Springer-Verlag, 1989

[14] R.L. Haupt and S.E. Haupt, "Practical Genetic Algorithms", Second Edition, Wiley Interscience Publication,1998.

[15] J.M. Smith, "Evolutionary Genetics", Second Edition, Oxford University Press Inc, 1998.

[16] J.Wen, S. Wang, S. Cheng, Q.H. Wu and D.W. Shimmin, "Measurement Based Power System Load Modeling Using A Population Diversity Genetic Algorithm", International Conference on Power System Technology, Vol. 1, 18-21 Aug 1998, pp. 771-775.

[17] S. Nootyaskool and B. Kruatrachue, "Hybrid Genetic Algorithm with Baum-Welch Algorithm by using Diversity Population Technique" International Symposium on Communications and Information Technologies 2006 (ISCIT2006), October 2006, pp. 15-20.

[18] Department of Information and Computer Science, University of California. "UCI Machine Learning Repository" [Online]. Available: http://www.ics.uci.edu/~mlearn/ MLRepository.html. 1998.

[19] K. Kangkan and B. Kruatrachue, "Minimal Consistent Subset Selection as Interger Nonlinear Programming Problem", International Symposium on Communications and Information Technologies 2006 (ISCIT2006), October 2006, pp. 54-58.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูลต่างๆ

การทดลองของ GAs กับชุดข้อมูลต่างๆ เป็นการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูลต่างๆ โดยมีการกำหนดค่าตัวแปรต่างๆ ดังนี้

- จำนวนประชากร (N_{pop}) กำหนดให้มีค่าเท่ากับ "40"
- ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) กำหนดให้มีค่าเท่ากับ "0.8"
- ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) กำหนดให้มีค่าเท่ากับ "0.5"
- ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) กำหนดให้มีค่าเท่ากับ "0.015"

ผลลัพธ์การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์ ของ GAs กับชุดข้อมูลต่างๆ แสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูลทั้ง 6 ชุด โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที, 30 นาที, 1 ชั่วโมง, 3 ชั่วโมง, 6 ชั่วโมง และ 12 ชั่วโมง และเนื่องจาก GAs และ HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละการทำงานจะให้ผลลัพธ์ไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

โดยจะแสดงเป็นตัวอย่างเพียงผลการทดลองของ GAs เพียง 2 ชุดข้อมูล ได้แก่ชุดข้อมูล IRIS และชุดข้อมูล PEN

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

I. ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS

ตารางที่ ก.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 1

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	11	00:02:10.58	10	00:00:58.70	8	00:02:12.76
30 นาที	10	00:16:19.86	9	00:16:20.68	6	00:23:39.95
1 ชั่วโมง	10	00:16:19.86	9	00:16:20.68	6	00:23:39.95
3 ชั่วโมง	10	00:16:19.86	9	00:16:20.68	6	00:23:39.95
6 ชั่วโมง	10	00:16:19.86	9	00:16:20.68	6	00:23:39.95
12 ชั่วโมง	10	00:16:19.86	9	00:16:20.68	6	00:23:39.95

ตารางที่ ก.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 2

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	13	00:00:25.82	12	00:0012.18	9	00:00:51.76
30 นาที	12	00:19:23.25	11	00:19:24.30	8	00:28:00.25
1 ชั่วโมง	12	00:19:23.25	11	00:19:24.30	8	00:28:00.25
3 ชั่วโมง	12	00:19:23.25	10	01:07:12.2	8	00:28:00.25
6 ชั่วโมง	10	04:28:33.36	9	04:28:33.63	6	04:28:33.36
12 ชั่วโมง	10	04:28:33.36	9	04:28:33.63	6	04:28:33.36

ตารางที่ ก.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 3

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	12	00:01:33.03	10	00:01:33.92	8	00:07:02.06
30 นาที	12	00:01:33.03	9	00:08:15.20	8	00:07:02.06
1 ชั่วโมง	12	00:01:33.03	9	00:08:15.20	8	00:07:02.06
3 ชั่วโมง	12	00:01:33.03	9	00:08:15.20	8	00:07:02.06
6 ชั่วโมง	12	00:01:33.03	9	00:08:15.20	8	00:07:02.06
12 ชั่วโมง	12	00:01:33.03	9	00:08:15.20	8	00:07:02.06

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 4

เวลา -	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา -
15 นาที	12	00:01:39.97	10	00:01:46.71	9	00:02:07.92
30 นาที	12	00:01:39.97	10	00:01:46.71	9	00:02:07.92
1 ชั่วโมง	12	00:01:39.97	10	00:01:46.71	9	00:02:07.92
3 ชั่วโมง	12	00:01:39.97	10	00:01:46.71	8	01:14:21.04
6 ชั่วโมง	12	00:01:39.97	10	00:01:46.71	8	01:14:21.04
12 ชั่วโมง	12	00:01:39.97	10	00:01:46.71	8	01:14:21.04

ตารางที่ ก.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล IRIS ครั้งที่ 5

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	12	00:02:56.80	10	00:03:23.94	9	00:03:23.94
30 นาที	12	00:02:56.80	10	00:03:23.94	8	00:27:05.27
1 ชั่วโมง	12	00:02:56.80	10	00:03:23.94	8	00:27:05.27
3 ชั่วโมง	12	00:02:56.80	10	00:03:23.94	8	00:27:05.27
6 ชั่วโมง	12	00:02:56.80	10	00:03:23.94	8	00:27:05.27
12 ชั่วโมง	12	00:02:56.80	10	00:03:23.94	8	00:27:05.27

II. ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN

ตารางที่ ก.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 1

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	4175	00:14:44.79	4147	00:14:51.10	4147	00:14:51.10
30 นาที	3936	00:27:26.63	3912	00:28:45.99	3912	00:28:45.99
1 ชั่วโมง	3747	00:54:38.31	3718	00:59:03.71	3718	00:59:03.71
3 ชั่วโมง	3493	02:58:38.08	3463	02:55:53.32	3463	02:55:53.32
6 ชั่วโมง	3314	05:37:52.53	3301	05:52:44.38	3301	05:52:44.38
12 ชั่วโมง	3194	11:35:34.26	3146	11:51:18.41	3146	11:51:18.41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.7 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 2

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวน -โปรโตไทป์	เวลา
15 นาที	4478	00:14:56.62	4442	00:14:49.87	4442	00:14:49.87
30 นาที	4203	00:29:58.62	4174	00:29:59.98	4174	00:29:59.98
1 ชั่วโมง	3998	00:58:17.10	3970	00:51:49.84	3970	00:51:49.84
3 ชั่วโมง	3672	02:57:18.83	3646	02:54:43.33	3646	02:54:43.33
6 ชั่วโมง	3469	05:53:42.63	3428	05:54:55.80	3428	05:54:55.80
12 ชั่วโมง	3327	11:56:51.80	3293	10:56:03.81	3293	10:56:03.81

ตารางที่ ก.8 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 3

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	4510	00:14:56.33	4470	00:14:56.33	4470	00:14:56.33
30 นาที	4268	00:29:54.75	4241	00:29:17.32	4241	00:29:17.32
1 ชั่วโมง	3978	00:58:57.64	3958	00:58:45.47	3958	00:58:45.47
3 ชั่วโมง	3658	02:58:12.16	3658	02:58:08.20	3658	02:58:08.20
6 ชั่วโมง	3522	05:57:58.11	3482	05:37:10.40	3482	05:37:10.40
12 ชั่วโมง	3333	11:56:11.12	3312	11:48:37.39	3312	11:48:37.39

ตารางที่ ก.9 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 4

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	4543	00:14:12.36	4510	00:14:32.57	4510	00:14:32.57
30 นาที	4295	00:29:52.66	4249	00:26:39.02	4249	00:26:39.02
1 ชั่วโมง	4003	00:58:35.27	3987	00:59:33.24	3987	00:59:33.24
3 ชั่วโมง	3697	02:57:44.99	3668	02:55:18.60	3668	02:55:18.60
6 ชั่วโมง	3506	05:41:01.14	3486	05:59:21.75	3486	05:59:21.75
12 ชั่วโมง	3353	10:55:57.51	3335	10:55:15.28	3335	10:55:15.28

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.10 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ GAs กับชุดข้อมูล PEN ครั้งที่ 5

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	4507	00:14:53.32	4504	00:14:32.23	4504	00:14:32.23
30 นาที	4224	00:29:23.13	4208	00:29:31.48	4208	00:29:31.48
1 ชั่วโมง	3971	00:52:11.22	3946	00:59:48.12	3946	00:59:48.12
3 ชั่วโมง	3647	02:52:13.60	3609	02:54:55.79	3609	02:54:55.79
6 ชั่วโมง	3609	05:57:37.86	3487	05:02:46.06	3487	05:02:46.06
12 ชั่วโมง	3385	11:26:22.33	3371	11:27:28.86	3371	11:27:28.86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูลต่างๆ

การทดลองของ HGAs กับชุดข้อมูลต่างๆ เป็นการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูลต่างๆ โดยมีการกำหนดค่าตัวแปรต่างๆ ดังนี้

- จำนวนประชากร (N_{pop}) กำหนดให้มีค่าเท่ากับ "10"
- ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) กำหนดให้มีค่าเท่ากับ "0.8"
- ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) กำหนดให้มีค่าเท่ากับ "0.5"
- ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) กำหนดให้มีค่าเท่ากับ "0.015"

ผลลัพธ์การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์ของ HGAs กับชุดข้อมูลต่างๆ แสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูลทั้ง 6 ชุด โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที, 30 นาที, 1 ชั่วโมง, 3 ชั่วโมง, 6 ชั่วโมง และ 12 ชั่วโมง และเนื่องจาก GAs และ HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละการทำงานจะให้ผลลัพธ์ไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

โดยจะแสดงเป็นตัวอย่างเพียงผลการทดลองของ GAs เพียง 2 ชุดข้อมูล ได้แก่ชุดข้อมูล IRIS และชุดข้อมูล PEN

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

I. ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS

ตารางที่ ข.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 1

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10	00:00:04.30	9	00:00:06.27	7	00:01:58.62
30 นาที	10	00:00:04.30	9	00:00:06.27	7	00:01:58.62
1 ชั่วโมง	10	00:00:04.30	9	00:00:06.27	7	00:01:58.62
3 ชั่วโมง	10	00:00:04.30	9	00:00:06.27	7	00:01:58.62
6 ชั่วโมง	10	00:00:04.30	9	00:00:06.27	7	00:01:58.62
12 ชั่วโมง	10	00:00:04.30	9	00:00:06.27	7	00:01:58.62

ตารางที่ ข.2 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 2

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10	00:00:04.38	9	00:0:08.59	7	00:01:20.57
30 นาที	10	00:00:04.38	9	00:0:08.59	7	00:01:20.57
1 ชั่วโมง	10	00:00:04.38	9	00:0:08.59	7	00:01:20.57
3 ชั่วโมง	10	00:00:04.38	9	00:0:08.59	7	00:01:20.57
6 ชั่วโมง	10	00:00:04.38	9	00:0:08.59	7	00:01:20.57
12 ชั่วโมง	10	00:00:04.38	9	00:0:08.59	7	00:01:20.57

ตารางที่ ข.3 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 3

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10	00:00:04.59	9	00:00:11.05	7	00:01:05.45
30 นาที	10	00:00:04.59	9	00:00:11.05	7	00:01:05.45
1 ชั่วโมง	10	00:00:04.59	9	00:00:11.05	7	00:01:05.45
3 ชั่วโมง	10	00:00:04.59	9	00:00:11.05	7	00:01:05.45
6 ชั่วโมง	10	00:00:04.59	9	00:00:11.05	7	00:01:05.45
12 ชั่วโมง	10	00:00:04.59	9	00:00:11.05	7	00:01:05.45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.4 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 4

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10	00:00:05.59	9	00:00:01.05	8	00:01:25.45
30 นาที	10	00:00:05.59	9	00:00:01.05	8	00:01:25.45
1 ชั่วโมง	10	00:00:05.59	9	00:00:01.05	7	00:30:50.74
3 ชั่วโมง	10	00:00:05.59	9	00:00:01.05	7	00:30:50.74
6 ชั่วโมง	10	00:00:05.59	9	00:00:01.05	7	00:30:50.74
12 ชั่วโมง	10	00:00:05.59	9	00:00:01.05	7	00:30:50.74

ตารางที่ ข.5 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล IRIS ครั้งที่ 5

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10	00:00:02.02	9	00:00:29.18	8	00:00:04.98
30 นาที	10	00:00:02.02	9	00:00:29.18	8	00:00:04.98
1 ชั่วโมง	10	00:00:02.02	9	00:00:29.18	7	00:35:58.40
3 ชั่วโมง	10	00:00:02.02	9	00:00:29.18	7	00:35:58.40
6 ชั่วโมง	10	00:00:02.02	9	00:00:29.18	7	00:35:58.40
12 ชั่วโมง	10	00:00:02.02	9	00:00:29.18	7	00:35:58.40

II. ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN

ตารางที่ ข.6 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 1

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
30 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
1 ชั่วโมง	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
3 ชั่วโมง	306	01:26:51.48	296	01:26:53.55	296	01:26:53.55
6 ชั่วโมง	306	01:26:51.48	296	01:26:53.55	296	01:26:53.55
12 ชั่วโมง	306	01:26:51.48	296	01:26:53.55	296	01:26:53.55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.7 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 2

เวลา -	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
30 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
1 ชั่วโมง	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
3 ชั่วโมง	307	01:25:12.43	283	02:50:24.86	283	02:50:24.86
6 ชั่วโมง	307	01:25:12.43	283	02:50:24.86	283	02:50:24.86
12 ชั่วโมง	307	01:25:12.43	283	02:50:24.86	283	02:50:24.86

ตารางที่ ข.8 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 3

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
30 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
1 ชั่วโมง	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
3 ชั่วโมง	319	01:36:31.71	287	01:09:25.52	287	01:09:25.52
6 ชั่วโมง	310	03:10:55.17	287	01:09:25.52	287	01:09:25.52
12 ชั่วโมง	310	03:10:55.17	287	01:09:25.52	287	01:09:25.52

ตารางที่ ข.9 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 4

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
30 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
1 ชั่วโมง	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
3 ชั่วโมง	314	01:25:56.62	274	01:07:57.30	274	01:07:57.30
6 ชั่วโมง	314	01:25:56.62	274	01:07:57.30	274	01:07:57.30
12 ชั่วโมง	314	01:25:56.62	274	01:07:57.30	274	01:07:57.30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.10 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs กับชุดข้อมูล PEN ครั้งที่ 5

เวลา	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา	จำนวนโปรโตไทป์	เวลา
15 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
30 นาที	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
1 ชั่วโมง	10092	00:00:00.00	10092	00:00:00.00	10092	00:00:00.00
3 ชั่วโมง	318	01:26:35.74	272	01:26:35.74	272	01:26:35.74
6 ชั่วโมง	318	01:26:35.74	272	01:26:35.74	272	01:26:35.74
12 ชั่วโมง	318	01:26:35.74	272	01:26:35.74	272	01:26:35.74

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

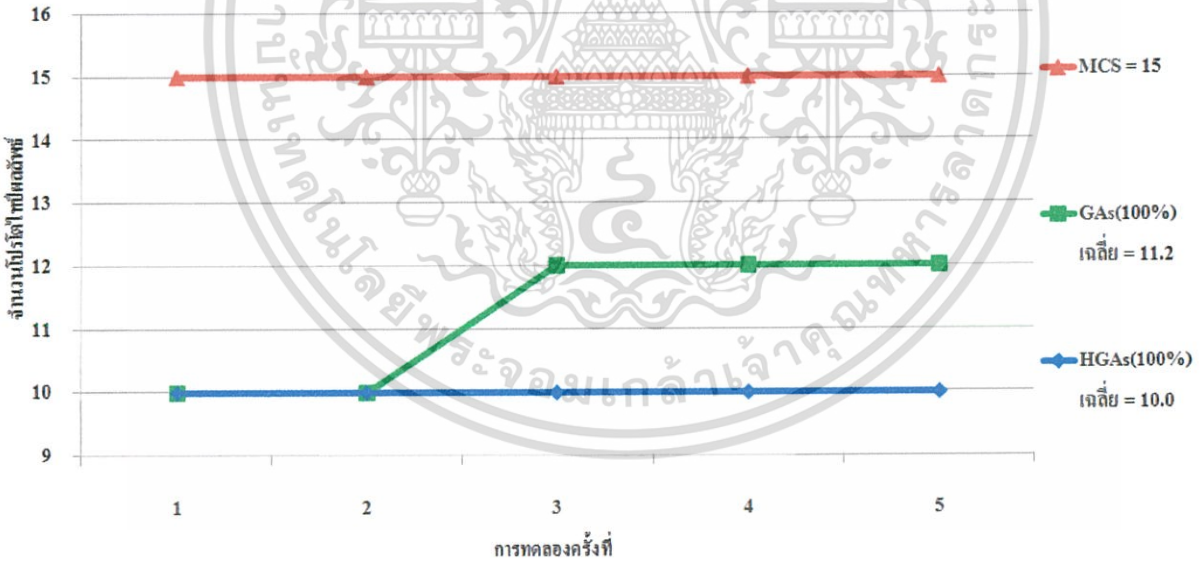
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค.

รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆ

รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกโปรโตไทป์วิธีต่างๆเป็นการแสดงรายละเอียดผลการทดลองทั้ง 5 ครั้งของวิธี MCS, GAs และ HGAs ที่แสดงในหัวข้อที่ 5.2.2 ในบทที่ 5แสดงเป็นแผนภูมิเส้นเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ของแต่ละวิธี โดยในผลลัพธ์ของวิธี MCS จะเหมือนกันทุกครั้ง เพราะวิธี MCS เป็นวิธีที่ตายตัวที่ให้ผลลัพธ์เหมือนกันทุกการทำงาน ส่วนผลลัพธ์ของ GAs และ HGAs จะแสดงเฉพาะกรณีสอดคล้อง 100% เท่านั้น เพื่อให้เหมาะสมกับการนำมาเปรียบเทียบกับผลลัพธ์ของวิธี MCS

I. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกโปรโตไทป์วิธีต่างๆ ของชุดข้อมูล IRIS

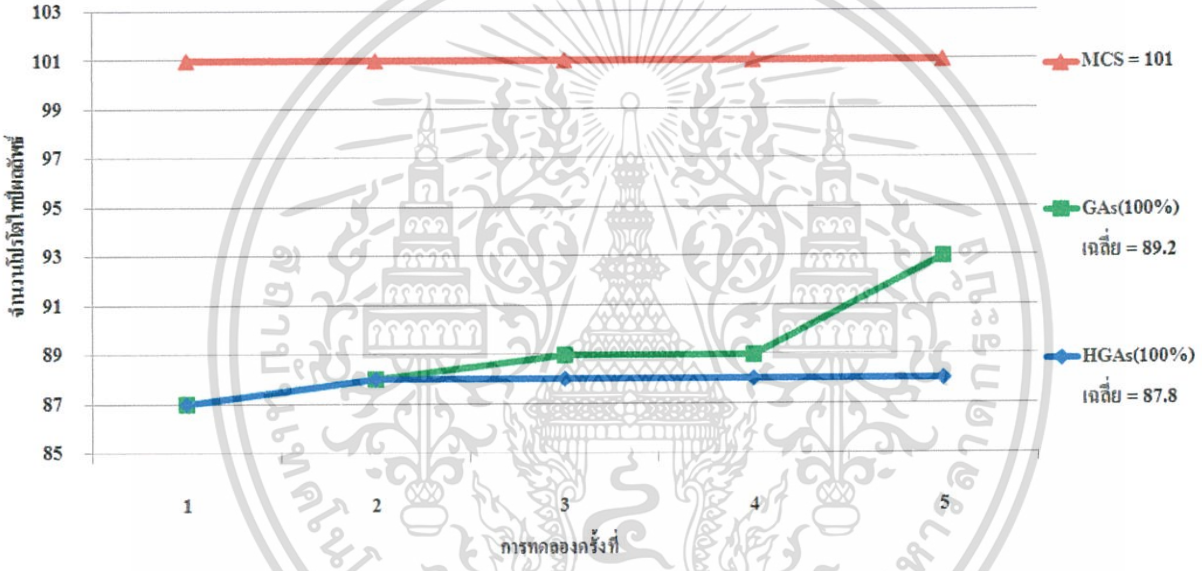


รูปที่ ค.1 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล IRIS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ ค.1 แสดงรายละเอียดการทดลองของวิธี MCS, GAs และ HGAs ของชุดข้อมูล IRIS ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์จำนวน 15 ตัว ส่วน GAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 11.2 ตัว และ HGAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 10.0 ตัว (จากจำนวนข้อมูลทั้งหมด 150 ตัว) ซึ่งจะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่า GAs และวิธี MCS ตามลำดับ และแม้ว่า GAs ให้ผลลัพธ์ที่มีจำนวนโปรโตไทป์เท่ากับ "10" ได้ แต่ก็หาได้เพียง 2 ครั้งจากการทดลอง 5 ครั้ง ในขณะที่ HGAs ให้ผลลัพธ์ที่มีจำนวนโปรโตไทป์เท่ากับ "10" ทั้ง 5 ครั้งที่ทำการทดลอง

II. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกโปรโตไทป์วิธีต่างๆ ของชุดข้อมูล ECOLI

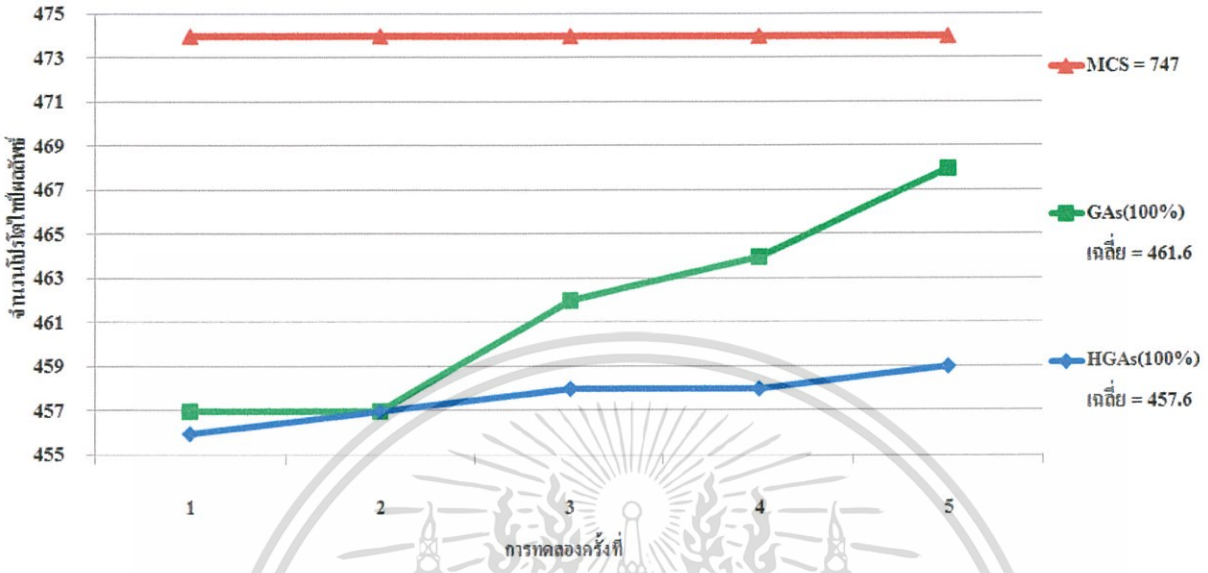


รูปที่ ค.2 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล ECOLI

จากรูปที่ ค.2 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs ของชุดข้อมูล ECOLI ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์จำนวน 101 ตัว ส่วน GAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 89.2 ตัว และ HGAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 87.8 ตัว (จากจำนวนข้อมูลทั้งหมด 336 ตัว) ซึ่งจะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่า GAs และวิธี MCS ตามลำดับ และแม้ว่าผลลัพธ์ของ GAs กับ HGAs จะใกล้เคียงกันแต่ผลลัพธ์ของ HGAs มีความสม่ำเสมอมากกว่า ในขณะที่ผลลัพธ์ของ GAs มีความเบี่ยงเบนค่อนข้างมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

III. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกโปรโตไทป์วิธีต่างๆ ของชุดข้อมูล CREDIT

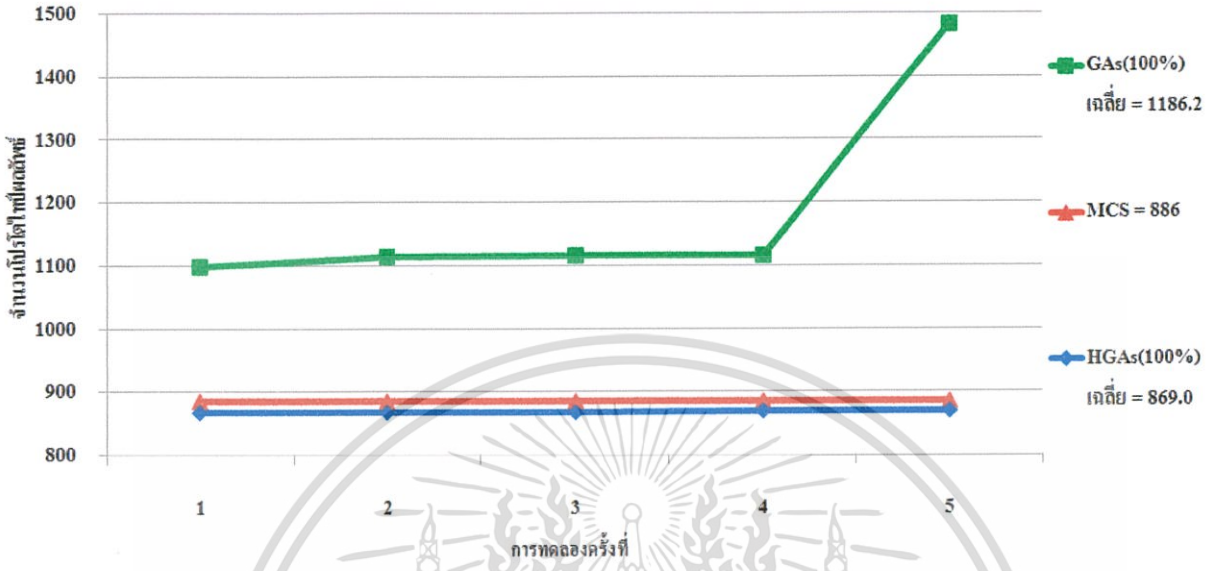


รูปที่ ค.3 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล CREDIT

จากรูปที่ ค.3 แสดงรายละเอียดการทดลองของวิธี MCS, GAs และ HGAs ของชุดข้อมูล CREDIT ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์จำนวน 747 ตัว ส่วน GAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 461.6 ตัว และ HGAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 457.6 ตัว (จากจำนวนข้อมูลทั้งหมด 1000 ตัว) ซึ่งจะเห็นว่า HGAs ให้ผลลัพธ์ที่ดีกว่า GAs และวิธี MCS ตามลำดับ แต่ผลลัพธ์ของ GAs แย่ลงอย่างเห็นได้ชัด มีการเบี่ยงเบนค่อนข้างมาก ในขณะที่ HGAs ยังให้ผลลัพธ์ที่ดีและค่อนข้างมีความสม่ำเสมอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

IV. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกโปรโตไทป์วิธีต่างๆ ของชุดข้อมูล YEST

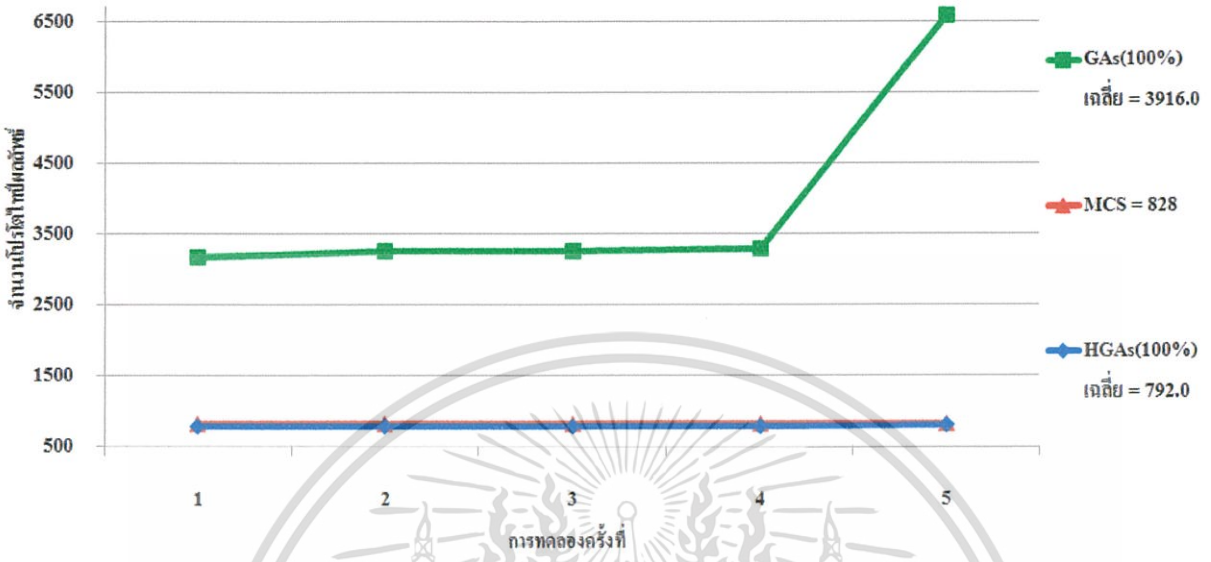


รูปที่ ค.4 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล YEAT

จากรูปที่ ค.4 แสดงรายละเอียดการทดลองของวิธี MCS, GAs และ HGAs ของชุดข้อมูล YEAT ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์จำนวน 886 ตัว ส่วน GAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 1186.2 ตัว และ HGAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 869.0 ตัว (จากจำนวนข้อมูลทั้งหมด 1484 ตัว) ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์ดีกว่า GAs ในขณะที่ HGAs ยังให้ผลลัพธ์ที่ดีกว่าทั้งวิธี MCS และ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

V. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกโปรโตไทป์วิธีต่างๆ ของชุดข้อมูล MUSK

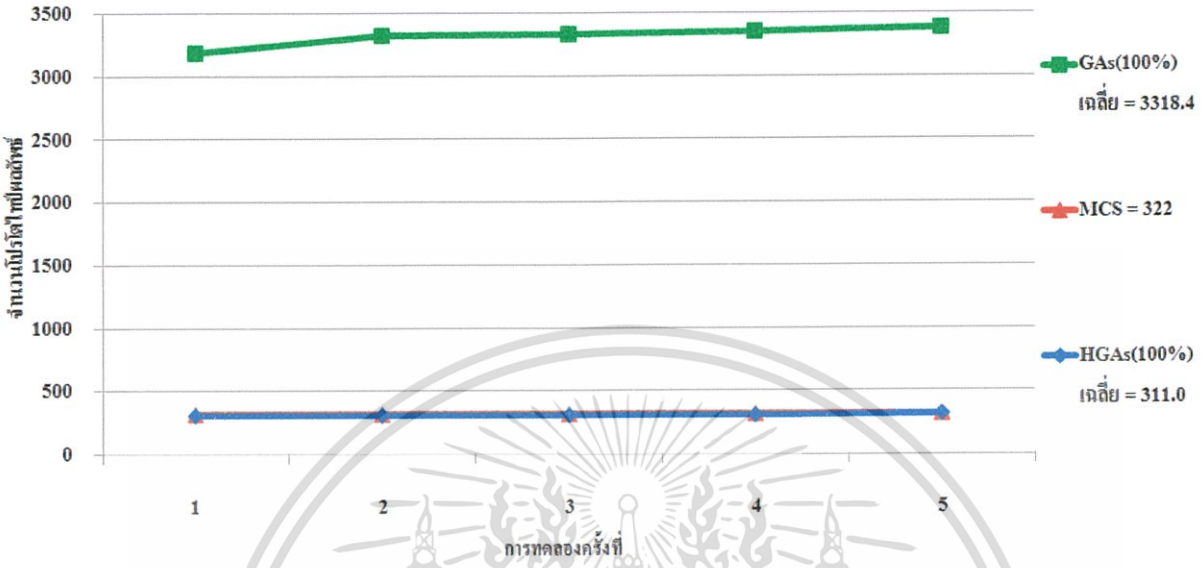


รูปที่ ค.5 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล MUSK

จากรูปที่ ค.5 แสดงรายละเอียดการทดลองของวิธี MCS, GAs และ HGAs ของชุดข้อมูล MUSK ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์จำนวน 828 ตัว ส่วน GAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 3916.0 ตัว และ HGAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 792.0 ตัว (จากจำนวนข้อมูลทั้งหมด 6598 ตัว) ซึ่งจะเห็นว่ายิ่งชุดข้อมูลมีข้อมูลจำนวนมากขึ้น GAs ก็จะให้ผลลัพธ์ที่ยิ่งแย่ลง ในขณะที่ HGAs ยังให้ผลลัพธ์ที่ดีกว่าทั้งวิธี MCS และ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

VI. รายละเอียดผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของของการเลือกโปรโตไทป์วิธีต่างๆ ของชุดข้อมูล PEN



รูปที่ ค.6 รายละเอียดการทดลองของวิธี MCS, GAs และ HGAs กับชุดข้อมูล PEN

จากรูปที่ ค.6 แสดงรายละเอียดการทดลองของวิธี MCS, GAs และ HGAs ของชุดข้อมูล PEN ซึ่งจะเห็นว่าวิธี MCS ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์จำนวน 322 ตัว ส่วน GAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 3318.4 ตัว และ HGAs ให้ผลลัพธ์เป็นโปรโตไทป์ผลลัพธ์เฉลี่ยจำนวน 331.0 ตัว (จากจำนวนข้อมูลทั้งหมด 10092 ตัว) ซึ่งจะเห็นว่า GAs ให้ผลลัพธ์ที่แย่กว่าวิธี MCS อย่างชัดเจน ในขณะที่ HGAs ยังให้ผลลัพธ์ที่ดีกว่าทั้งวิธี MCS และ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ง.

ตัวอย่างการทดลองของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้

ตัวอย่างการทดลองของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ เป็นการทดลองของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ ที่อ้างอิงตามบทความที่มีชื่อว่า "Minimal Consistent Subset Selection as Integer Nonlinear Programming Problem" [19] ที่ประยุกต์เอาหลักการทางคณิตศาสตร์เข้ามาแก้ปัญหาการเลือกโปรโตไทป์ ซึ่งทำให้ทราบว่าผลลัพธ์ที่ดีที่สุดของชุดข้อมูล IRIS คือ โปรโตไทป์ผลลัพธ์ที่มีจำนวนโปรโตไทป์เท่ากับ "10" ตัว (โดยโปรโตไทป์ผลลัพธ์จะต้องสอดคล้องกับชุดข้อมูลตั้งต้นด้วย) จึงทำการแสดงผลการทดลอง เพื่อแสดงให้เห็นถึงความสำคัญของกระบวนการต่างๆ ใน HGAs ที่ต้องทำงานร่วม จนทำให้ได้ผลลัพธ์ที่ดีที่สุด

ผลการทดลองของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ มี 2 ผลการทดลอง ดังนี้

1. ตัวอย่างการทดลองที่ 1 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้
2. ตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้

โดยผลตัวอย่างการทดลองที่ 1 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดแสดงในบทที่ 5 ส่วนผลตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้ แสดงในภาคผนวกนี้

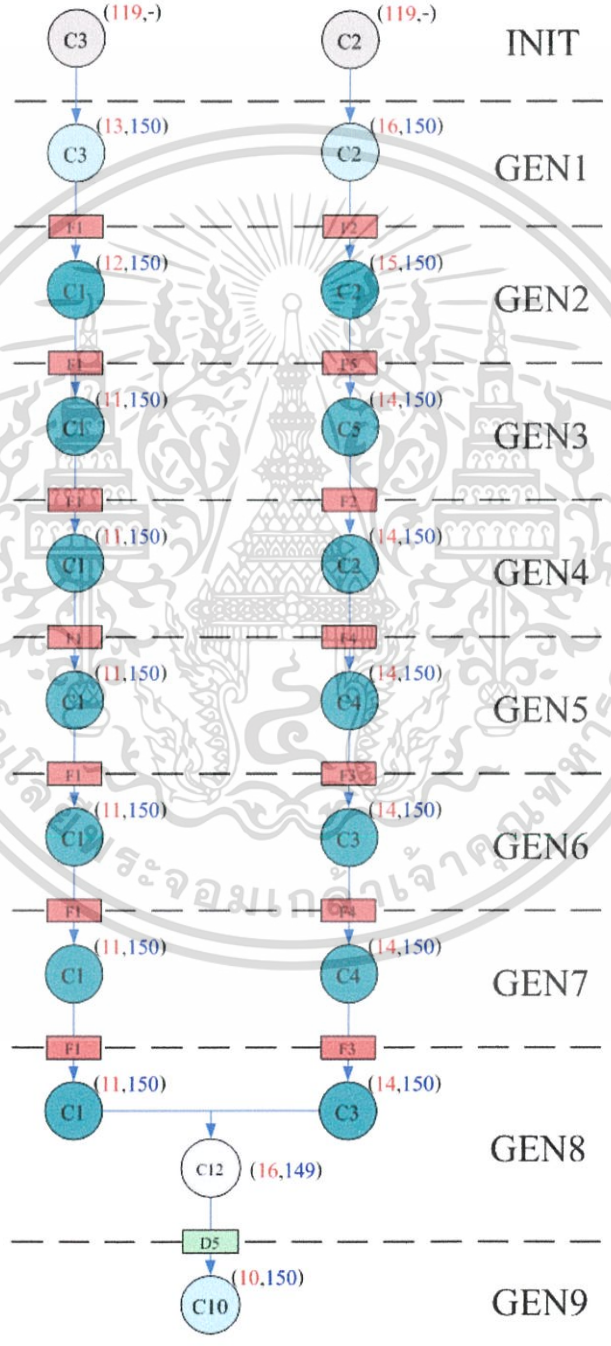
HGAs ที่ใช้ในการทดลองมีขั้นตอนการทำงานอ้างอิงจากหัวข้อที่ 4.2 ในบทที่ 4 และมีการกำหนดค่าตัวแปรต่างๆ ดังนี้

- จำนวนประชากร (N_{pop}) กำหนดให้มีค่าเท่ากับ "10"
- ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini}) กำหนดให้มีค่าเท่ากับ "0.8"
- ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c) กำหนดให้มีค่าเท่ากับ "0.5"
- ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m) กำหนดให้มีค่าเท่ากับ "0.015"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

I. รายละเอียดผลตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้

ผลตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS ที่สามารถหาผลลัพธ์ที่ดีที่สุดได้เป็นผลการทดลองที่ HGAs หาผลลัพธ์ที่มีจำนวนโปรโตไทป์เท่ากับ “10” ได้ในเจเนอเรชั่นที่ “9” โดยจะแสดงในลักษณะโครงสร้างครอบครัวเพื่ออธิบายถึงที่มาของผลลัพธ์ที่ดีที่สุด



รูปที่ ง.1 โครงสร้างครอบครัวของตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ ง.1 เป็นการแสดงโครงสร้างครอบครัวของตัวอย่างการทดลองที่ 2 ของ HGAs กับชุดข้อมูล IRIS โดยแสดงการพัฒนาของโครโมโซมตั้งแต่เจเนอเรชั่นที่ 1 ไปจนถึงเจเนอเรชั่นที่ 9 ซึ่งเมื่อทำการพิจารณาโครโมโซมบรรพบุรุษของโครโมโซมที่ดีที่สุดนี้ก็จะสามารถสรุปข้อสังเกตได้ 3 ข้อดังนี้

1. จะสังเกตเห็นว่าโครโมโซมบรรพบุรุษของโครโมโซมที่เป็นผลลัพธ์ที่ดีที่สุดมีเพียง 2 ตัวเท่านั้น ซึ่งก็คือโครโมโซมลำดับที่ 2 และโครโมโซมลำดับที่ 3 ในเจเนอเรชั่นที่ 1 ที่มีการพัฒนาจากเจเนอเรชั่นที่ 1 ไปยังเจเนอเรชั่นที่ 3 แต่ตั้งแต่เจเนอเรชั่นที่ 3 ไปจนถึงเจเนอเรชั่นที่ 8 ไม่มีการพัฒนาต่อเลยซึ่งนี่เองเป็นข้อด้อยของวิธี MCS ที่ให้ผลลัพธ์ติดอยู่ใน “Local Optimum” จนกระทั่งโครโมโซมทั้ง 2 ผ่านกระบวนการทางพันธุกรรมในเจเนอเรชั่นที่ 8 ทำให้ได้โครโมโซมลูกลำดับที่ 12 ซึ่งจะพัฒนาไปเป็นโครโมโซมที่ดีที่สุดต่อไป หมายความว่ากระบวนการทางพันธุกรรมเป็นสิ่งสำคัญที่นำไปสู่ผลลัพธ์ที่ดีที่สุด

2. จะสังเกตเห็นว่าโครโมโซมลำดับที่ 12 ในเจเนอเรชั่นที่ 8 ถูกเลือกไปเป็นประชากรในรุ่นถัดไปโดย “Diversity Selection” หมายความว่าการใช้เทคนิค “Diversity” เป็นสิ่งสำคัญที่นำไปสู่ผลลัพธ์ที่ดีที่สุด (โดยเฉพาะเทคนิค “Diversity Selection”)

3. จะสังเกตเห็นว่าโครโมโซมลำดับที่ 10 ในเจเนอเรชั่นที่ 9 ซึ่งเป็นผลลัพธ์ที่ดีที่สุดมาจากการพัฒนาจากโครโมโซมลำดับที่ 12 ในเจเนอเรชั่นที่ 8 โดยตรง นั่นหมายความว่าในการค้นหาคำตอบของ HGAs นั้น แม้ว่า GAs จะเป็นตัวช่วยกระจายค้นหาคำตอบในบริเวณกว้าง แต่คำตอบที่ได้ก็จะมาจากกระบวนการ MCS ที่เป็นตัวกรองคำตอบที่ได้นั้นอีกทีนั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก จ.

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อย

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อย เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จาก HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย กับ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมด

ขั้นตอนการทำงานของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการกำหนดค่าตัวแปรต่างๆ เช่นเดียวกับหัวข้อที่ 5.2.2

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมด ก็จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 และมีการกำหนดค่าตัวแปรต่างๆ เช่นเดียวกับหัวข้อที่ 5.2.2 เช่นกัน แต่จะแตกต่างกันตรงที่จะใช้ประบวนการ MCS แบบที่เลือกโปรโตไทป์จากข้อมูลทั้งหมดในชุดข้อมูลกล่าวคือโปรโตไทป์ผลลัพธ์ที่ได้อาจจะไม่เป็นเซตย่อยของ Candidate Set ก็ได้

โดยผลลัพธ์จะแสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูล IRIS, ชุดข้อมูล ECOLI, ชุดข้อมูล CREDIT และชุดข้อมูล YEAST โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที และ 30 นาที และเนื่องจาก HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละครั้งจะไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ SS คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ จ.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อยของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย

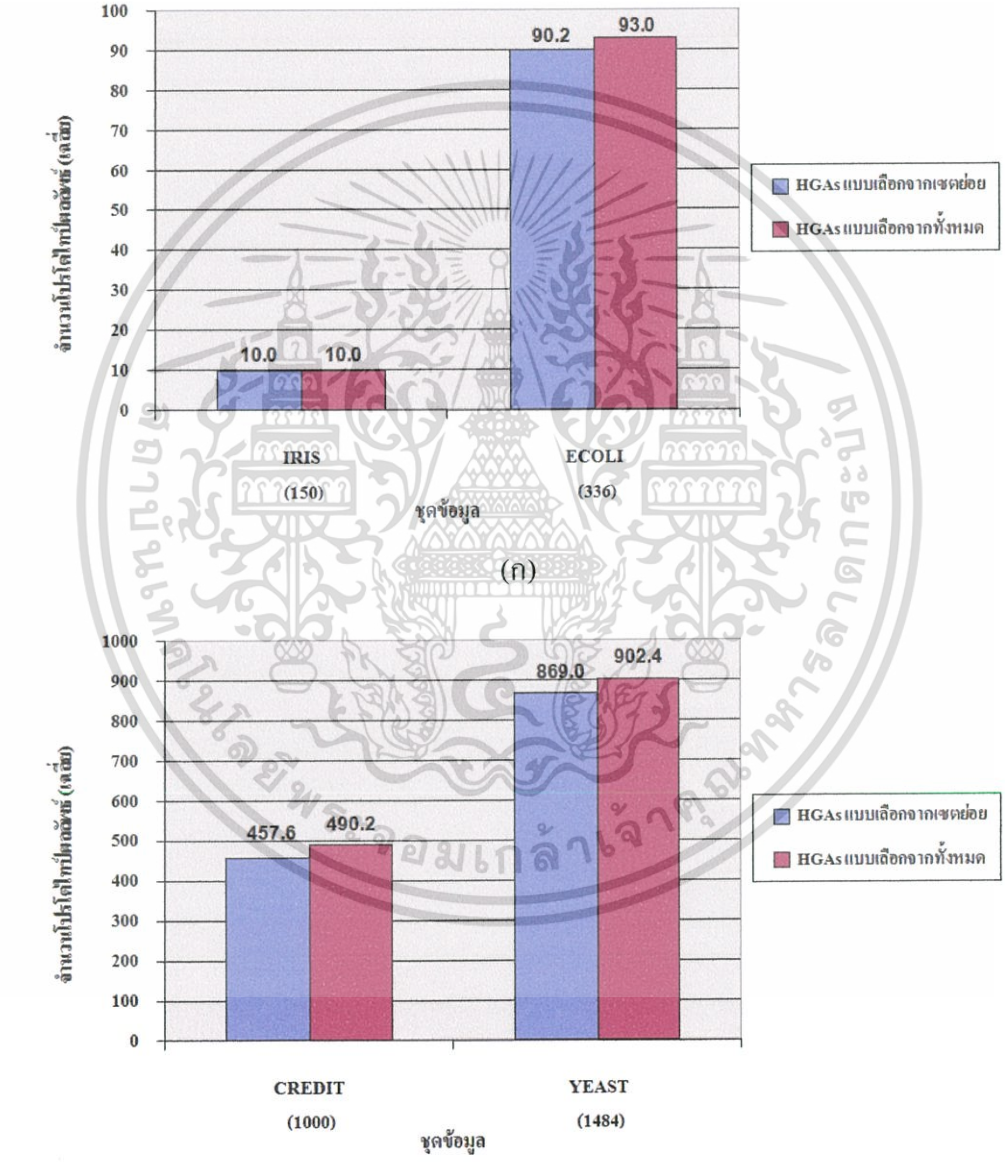
ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:04.90	9.0	00:00:06.22	7.4	00:01:32.39
ECOLI (336)	90.2	00:19:51.51	86.2	00:21:45.62	84.6	00:21:04.57
CREDIT (1000)	457.6	00:01:33.25	457.8	00:01:33.25	440.0	00:01:49.90
YEAST (1484)	869.0	00:07:07.28	869.0	00:07:07.28	853.2	00:05:03.13

ตารางที่ จ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อยของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมด

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:23.56	10.0	00:00:23.53	9.6	00:00:49.07
ECOLI (336)	93.0	00:05:27.84	89.8	00:13:37.93	87.6	00:06:05.16
CREDIT (1000)	490.2	00:00:30.20	441.2	00:17:16.79	437.6	00:19:59.04
YEAST (1484)	902.4	00:02:02.61	858.8	00:18:51.30	851.0	00:12:17.94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ จ.1 และตารางที่ จ.2 เป็นตารางผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อยของ HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อย และHGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมดของทั้ง 4 ชุดข้อมูล (ตามลำดับ) โดยช่องจำนวนโปรโตไทป์ และเวลาแสดงจำนวนโปรโตไทป์ผลลัพธ์ และเวลาการทำงาน (ตามลำดับ) โดยทั้งจำนวนโปรโตไทป์ และเวลาการทำงานจะพิจารณาจากค่าเฉลี่ยจากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 30 นาที)



(ก)

(ข)

รูปที่ จ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อย (เฉพาะกรณีสอดคล้อง 100%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ จ.1 เป็นการแสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของการเลือกจากเซตย่อย (เฉพาะกรณีสอดคล้อง 100%) โดยรูป (ก) แสดงผลการทดลองชุดข้อมูลขนาดเล็ก และรูป (ข) แสดงผลการทดลองชุดข้อมูลขนาดกลาง โดยแนวแกนตั้งแสดงถึงจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ยที่ได้จากแต่ละวิธี ส่วนแนวแกนนอนแสดงถึงชุดข้อมูลที่ใช้ในการทดลอง และตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

จากตารางที่ จ.1 และตารางที่ จ.2 เมื่อพิจารณาในกรณีความสอดคล้อง 100% จะเห็นว่า HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อยให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมดในทุกชุดข้อมูล (พิจารณาได้จากรูปที่ ค.1)

ส่วนเมื่อพิจารณาในกรณีความสอดคล้อง 99% และกรณีความสอดคล้อง 98% จะเห็นว่า HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อยจะให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมดเฉพาะชุดข้อมูล IRIS และชุดข้อมูล ECOLI เท่านั้น ทั้งนี้เพราะชุดข้อมูล IRIS และชุดข้อมูล ECOLI เป็นชุดข้อมูลที่มีจำนวนข้อมูลไม่มากนัก การเลือกโปรโตไทป์จาก Candidate Set ยังให้ผลลัพธ์ที่ดีกว่า แต่ในชุดข้อมูล CREDIT และชุดข้อมูล YEAST ซึ่งเป็นชุดข้อมูลที่มีจำนวนข้อมูลมากขึ้น การเลือกโปรโตไทป์จาก Candidate Set ไม่เพียงพอที่จะทำให้ได้ผลลัพธ์ที่ดี ดังนั้นการเลือกโปรโตไทป์จากข้อมูลทั้งหมดในชุดข้อมูลจึงให้ผลลัพธ์ที่ดีกว่า

อย่างไรก็ตามสิ่งสำคัญในการเลือกโปรโตไทป์ คือ โปรโตไทป์ผลลัพธ์จะต้องมีความสอดคล้องกับชุดข้อมูลตั้งต้น 100% ซึ่งจะเห็นได้ว่าในกรณีความสอดคล้อง 100% นั้น HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากเซตย่อยให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้กระบวนการ MCS แบบที่เลือกจากทั้งหมดในทุกชุดข้อมูล ดังนั้นจึงสรุปได้ว่าการใช้กระบวนการ MCS แบบเลือกจากเซตย่อยให้ผลลัพธ์ที่ดีกว่ากระบวนการ MCS แบบเลือกจากทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ฉ.

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton”

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จาก HGAs แบบที่ใช้เทคนิค “Singleton” กับ HGAs แบบที่ไม่ใช้เทคนิค “Singleton” เพื่อแสดงให้เห็นว่าการใช้เทคนิค “Singleton” เข้ามาร่วมทำงานให้ผลลัพธ์ที่ดีกว่าการไม่ใช้เทคนิค “Singleton”

ขั้นตอนการทำงานของ HGAs แบบที่ใช้เทคนิค “Singleton” จะอ้างอิงหัวข้อที่ 4.2 ในบทที่ 4 กล่าวคือมีการนำเทคนิค “Singleton” เข้ามาทำงานร่วมด้วย และมีการกำหนดค่าตัวแปรต่างๆ เหมือนกับหัวข้อที่ 5.2.2

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ไม่ใช้เทคนิค “Singleton” จะอ้างอิงหัวข้อที่ 4.2 ในบทที่ 4 และมีการกำหนดตัวแปรต่างๆ เช่นเดียวกับหัวข้อที่ 5.2.2 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำเทคนิค “Singleton” เข้ามาทำงานร่วมด้วย

โดยผลลัพธ์จะแสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูล IRIS, ชุดข้อมูล ECOLI, ชุดข้อมูล CREDIT และชุดข้อมูล YEAST โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที และ 30 นาที และเนื่องจาก HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละครั้งจะไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ฉ.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Singleton" ของ HGAs แบบที่ใช้เทคนิค "Singleton"

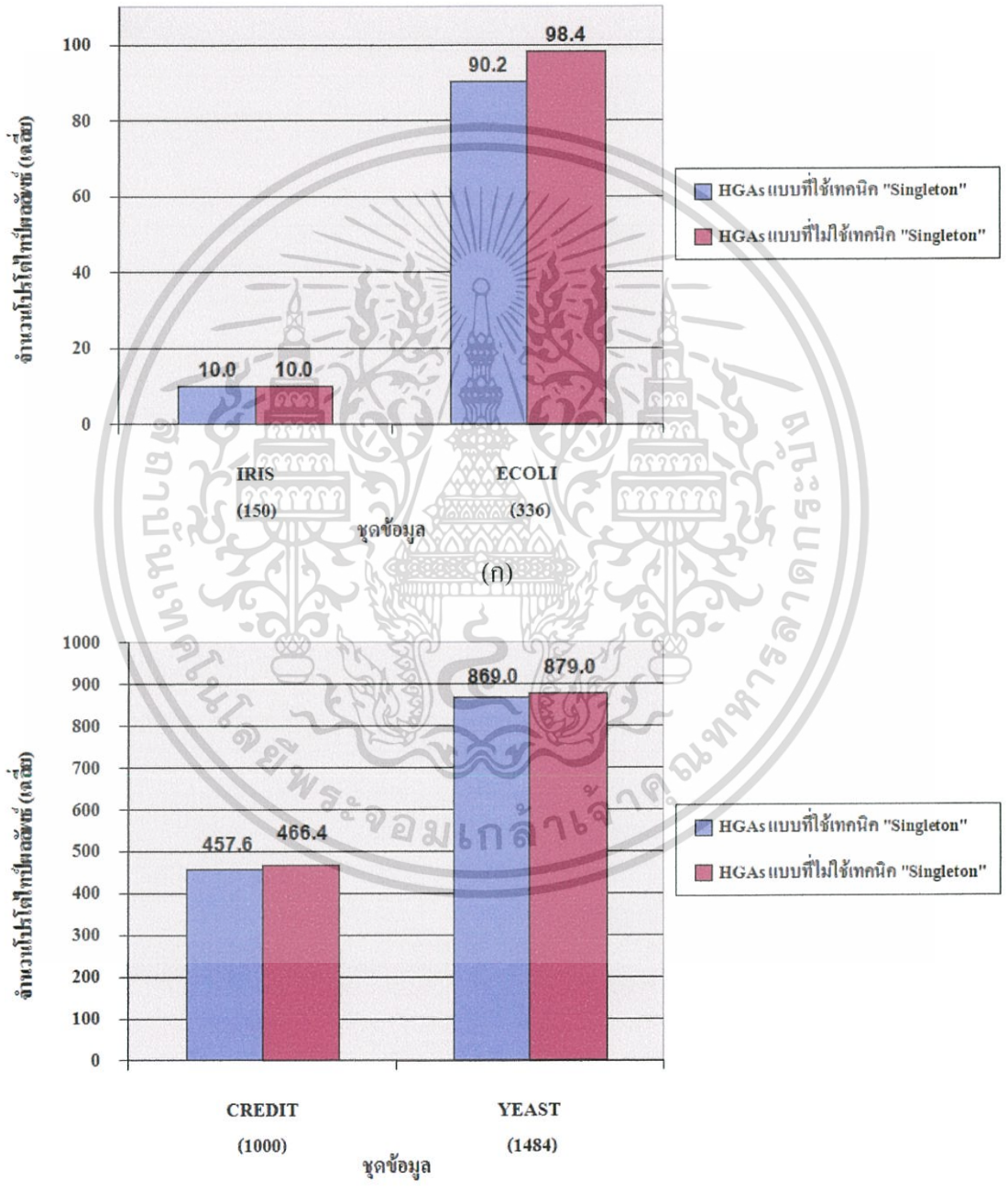
ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:04.90	9.0	00:00:06.22	7.4	00:01:32.39
ECOLI (336)	90.2	00:19:51.51	86.2	00:21:45.62	84.6	00:21:04.57
CREDIT (1000)	457.6	00:01:33.25	457.8	00:01:33.25	440.0	00:01:49.90
YEAST (1484)	869.0	00:07:07.28	869.0	00:07:07.28	853.2	00:05:03.13

ตารางที่ ฉ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Singleton" ของ HGAs แบบที่ไม่ใช้เทคนิค "Singleton"

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:26.74	7.6	00:06:12.35	3.0	00:00:55.12
ECOLI (336)	98.4	00:00:05.83	90.6	00:18:53.96	79.6	00:18:46.67
CREDIT (1000)	466.4	00:04:01.19	466.4	00:04:01.19	466.4	00:04:01.19
YEAST (1484)	879.0	00:22:43.85	879.0	00:22:43.85	879.0	00:22:43.85

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ฉ.1 และตารางที่ ฉ.2 เป็นตารางผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” ของ HGAs แบบที่ใช้เทคนิค “Singleton” และ HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ของทั้ง 4 ชุดข้อมูล (ตามลำดับ) โดยช่องจำนวนโปรโตไทป์ และเวลาแสดงจำนวนโปรโตไทป์ผลลัพธ์ และเวลาการทำงาน (ตามลำดับ) โดยทั้งจำนวนโปรโตไทป์ และเวลาการทำงานจะพิจารณาจากค่าเฉลี่ยจากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 30 นาที)



รูปที่ ฉ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” (เฉพาะกรณีสอดคล้อง 100%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ฉ.1 เป็นการแสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Singleton” (เฉพาะกรณีสอดคล้อง 100%) โดยรูป (ก) แสดงผลการทดลองชุดข้อมูลขนาดเล็ก และรูป (ข) แสดงผลการทดลองชุดข้อมูลขนาดกลาง โดยแนวแกนตั้งแสดงถึงจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ยที่ได้จากแต่ละวิธี ส่วนแนวแกนนอนแสดงถึงชุดข้อมูลที่ใช้ในการทดลอง และตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

จากตารางที่ ฉ.1 และตารางที่ ฉ.2 จะเห็นว่า HGAs แบบที่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ในทุกกรณีความสอดคล้อง เฉพาะชุดข้อมูล CREDIT และชุดข้อมูล YEAST ส่วนในชุดข้อมูล IRIS นั้น HGAs แบบที่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” เฉพาะกรณีความสอดคล้อง 100% เท่านั้น ส่วนในชุดข้อมูล ECOLI นั้น HGAs แบบที่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ทั้งในกรณีความสอดคล้อง 100% และกรณีความสอดคล้อง 98%

จากผลการทดลองข้างต้น สรุปได้ว่าในชุดข้อมูลที่มีข้อมูลจำนวนไม่มาก อย่างเช่น ชุดข้อมูล IRIS จะเห็นว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้เทคนิค “Singleton” เพราะชุดข้อมูล IRIS มีจำนวนข้อมูลไม่มากนัก ทำให้ HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ยังสามารถเฟ้นสุ่มหาคำตอบที่ดีได้ ประกอบกับเทคนิค “Singleton” นั้นจะทำให้ออบเจ็คที่ไม่มีออบเจ็คอื่นสามารถครอบคลุมได้ถูกเลือกเป็นโปรโตไทป์ ซึ่งเป็นสิ่งที่ดีในกรณีความสอดคล้อง 100% แต่ในกรณีความสอดคล้อง 99% และความสอดคล้อง 98% นั้นอาจไม่จำเป็นจะต้องเลือก Singleton เป็นโปรโตไทป์เพราะไม่จำเป็นต้องจำแนกข้อมูลในชุดข้อมูลให้ถูกต้องทั้งหมด และเมื่อพิจารณาชุดข้อมูลที่มีข้อมูลจำนวนมากขึ้น อย่างเช่น ชุดข้อมูล ECOLI จะเห็นว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้เทคนิค “Singleton” แค่ในกรณีความสอดคล้อง 98% เท่านั้น และเมื่อพิจาณาต่อไปในชุดข้อมูลที่มีข้อมูลจำนวนมากขึ้น อย่างเช่น ชุดข้อมูล CREDIT และชุดข้อมูล YEAST เห็นว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ไม่สามารถให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ใช้เทคนิค “Singleton” เลย เพราะเมื่อชุดข้อมูลมีจำนวนข้อมูลมากขึ้น ก็จะทำให้ปริภูมิคำตอบมีขนาดเพิ่มมากขึ้น ทำให้ HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ไม่สามารถเฟ้นสุ่มหาคำตอบที่ดีได้เลยในทุกกรณีความสอดคล้อง

อย่างไรก็ตามในการเลือกโปรโตไทป์ที่ดีนั้นโปรโตไทป์ผลลัพธ์ควรจะต้องมีความสอดคล้องกับชุดข้อมูลตั้งต้น เพื่อยืนยันว่าประสิทธิภาพการจำแนกข้อมูลที่มีโปรโตไทป์ผลลัพธ์เป็นข้อมูลอ้างอิงมีความใกล้เคียงกันประสิทธิภาพการจำแนกข้อมูลที่มีชุดข้อมูลเป็นข้อมูลอ้างอิง และจากผลการทดลองข้างต้นก็จะเห็นว่าในกรณีความสอดคล้อง 100% นั้น HGAs แบบที่ใช้เทคนิค “Singleton” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Singleton” ในทุกชุดข้อมูล (พิจารณาได้จากรูปที่ ฉ.1) ดังนั้นจึงสรุปได้ว่าการนำเทคนิค “Singleton” เข้ามาทำงานร่วมกับ HGAs ทำให้ได้ผลลัพธ์ที่ดีขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ช.

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity"

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity" เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จาก HGAs แบบที่ใช้เทคนิค "Diversity" กับ HGAs แบบที่ไม่ใช้เทคนิค "Diversity" เพื่อแสดงให้เห็นว่าการใช้เทคนิค "Diversity" เข้ามาร่วมทำงานให้ผลลัพธ์ที่ดีกว่าการไม่ใช้เทคนิค "Diversity"

การทดลองของเทคนิค "Diversity" ที่นำมาใช้ใน HGAs แบ่งได้ 4 แบบ ดังนี้

1. HGAs แบบที่ใช้ทั้งเทคนิค "Diversity Crossover" และ "Diversity Selection"
2. HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Crossover"
3. HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Selection"
4. HGAs แบบที่ไม่ใช้ทั้งเทคนิค "Diversity Crossover" และ "Diversity Selection"

ขั้นตอนการทำงานของ HGAs แบบที่ใช้ทั้งเทคนิค "Diversity Crossover" และ "Diversity Selection" จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton", เทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วย

ส่วนขั้นตอนกาทำงานของ HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Crossover" จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำเทคนิค "Diversity Crossover" เข้ามาทำงานร่วมด้วย กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton" และเทคนิค "Diversity Crossover" เข้ามาทำงานร่วมด้วยเท่านั้น

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Selection" จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วย กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วยเท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ไม่ใช้ทั้งเทคนิค "Diversity Crossover" และ "Diversity Selection" จะอ้างอิงหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน แต่จะแตกต่างที่ไม่ได้นำทั้งเทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วย กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton" เข้ามาทำงานร่วมด้วยเท่านั้น

โดยผลลัพธ์จะแสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูล IRIS, ชุดข้อมูล ECOLI, ชุดข้อมูล CREDIT และชุดข้อมูล YEAST โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที และ 30 นาที และเนื่องจาก HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละครั้งจะไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

ารางที่ ช.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity" ของ HGAs แบบที่ใช้ทั้งเทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection"

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:04.90	9.0	00:00:06.22	7.4	00:01:32.39
ECOLI (336)	90.2	00:19:51.51	86.2	00:21:45.62	84.6	00:21:04.57
CREDIT (1000)	457.6	00:01:33.25	457.8	00:01:33.25	440.0	00:01:49.90
YEAST (1484)	869.0	00:07:07.28	869.0	00:07:07.28	853.2	00:05:03.13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ช.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity" ของ HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Crossover"

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:33.50	9.0	00:01:13.14	7.8	00:04:00.05
ECOLI (336)	96.2	00:04:17.63	92.8	00:00:56.68	91.0	00:08:43.32
CREDIT (1000)	460.8	00:01:25.63	459.2	00:01:25.63	450.4	00:01:02.23
YEAST (1484)	871.0	00:05:29.05	871.0	00:05:29.05	853.6	00:04:45.59

ตารางที่ ช.3 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity" ของ HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Selection"

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:09.35	9.0	00:00:09.55	7.8	00:01:02.19
ECOLI (336)	91.0	00:21:37.25	86.4	00:25:27.05	85.0	00:24:50.14
CREDIT (1000)	460.6	00:02:06.30	458.8	00:02:06.30	448.2	00:01:53.07
YEAST (1484)	869.8	00:07:27.93	869.8	00:07:27.93	853.4	00:07:15.32

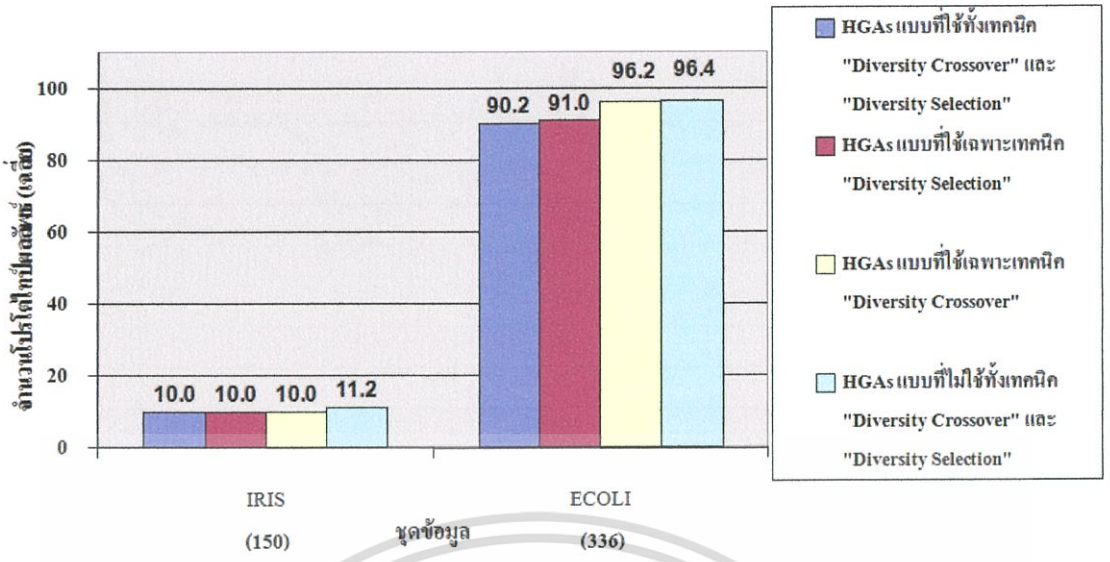
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ช.4 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity" ของ HGAs แบบที่ไม่ใช้ทั้งเทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection"

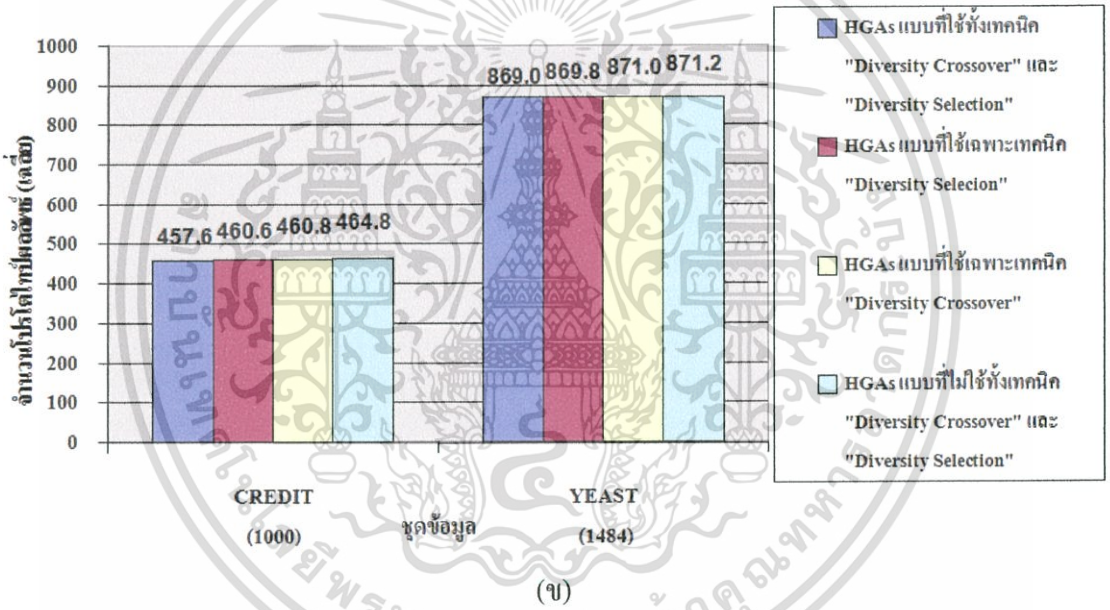
ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	11.2	00:12:34.37	9.6	00:15:55.71	8.6	00:16:04.80
ECOLI (336)	96.4	00:07:29.88	93.0	00:15:39.38	91.0	00:16:50.94
CREDIT (1000)	464.8	00:02:27.21	454.4	00:14:04.50	444.4	00:11:42.58
YEAST (1484)	871.2	00:05:10.16	871.2	00:05:10.16	854.8	00:05:11.77

ตารางที่ ช.1 , ตารางที่ ช.2, ตารางที่ ช.3 และตารางที่ ช.4 เป็นตารางผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค "Diversity" ของ HGAs แบบที่ใช้ทั้งเทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection", HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Crossover", HGAs แบบที่ใช้เฉพาะเทคนิค "Diversity Selection" และ HGAs แบบที่ไม่ใช้ทั้งเทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" ของทั้ง 4 ชุดข้อมูล (ตามลำดับ) โดยช่องจำนวนโปรโตไทป์ และเวลา แสดงจำนวนโปรโตไทป์ผลลัพธ์ และเวลาการทำงาน (ตามลำดับ) โดยทั้งจำนวนโปรโตไทป์ และเวลาการทำงานจะพิจารณาจากค่าเฉลี่ยจากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 30 นาที)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



(ข)

รูปที่ ช.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” (เฉพาะกรณีสอดคล้อง 100%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ช.1 เป็นการแสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของเทคนิค “Diversity” (เฉพาะกรณีสอดคล้อง 100%) โดยรูป (ก) แสดงผลการทดลองชุดข้อมูลขนาดเล็ก และรูป (ข) แสดงผลการทดลองชุดข้อมูลขนาดกลาง โดยแนวแกนตั้งแสดงถึงจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ยที่ได้จากแต่ละวิธี ส่วนแนวแกนนอนแสดงถึงชุดข้อมูลที่ใช้ในการทดลอง และตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

จากตารางที่ ช.1 ตารางที่ ช.2 ตารางที่ ช.3 และตารางที่ ช.4 จะเห็นว่า HGAs แบบที่ใช้เทคนิค “Diversity” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Diversity” ทุกแบบทุกกรณีความสอดคล้อง ทุกชุดข้อมูล (พิจารณาได้จากรูปที่ ช.1)

เมื่อพิจารณาเปรียบเทียบกันระหว่าง “Diversity Crossover” กับ “Diversity Selection” จะเห็นว่า HGAs แบบที่ใช้เฉพาะเทคนิค “Diversity Selection” ให้ผลลัพธ์ที่ดีกว่า HGAs แบบที่ไม่ใช้เทคนิค “Diversity Selection” และ HGAs แบบที่ไม่ใช้ทั้งเทคนิค “Diversity Selection” และเทคนิค “Diversity Selection” ตามลำดับ หมายความว่าเทคนิค “Diversity Selection” มีผลทำให้ HGAs ส่วนมากให้ผลลัพธ์ที่ดีมากกว่าเทคนิค “Diversity Crossover” และเมื่อนำมาใช้ร่วมกันก็จะทำให้ HGAs ให้ผลลัพธ์ที่ดีมากกว่าใช้เพียงเทคนิคใดเทคนิคหนึ่ง ดังนั้นจึงสรุปได้ว่าการนำเทคนิค “Diversity” เข้ามาทำงานร่วมกับ HGAs ทำให้ได้ผลลัพธ์ที่ดีขึ้น

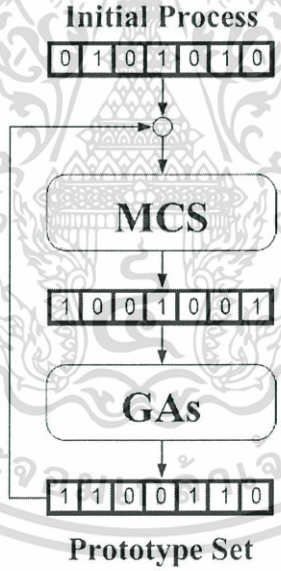
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ซ.

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของลำดับการทำงาน

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของลำดับการทำงาน เป็นการทดลองเพื่อเปรียบเทียบจำนวนโปรโตไทป์ผลลัพธ์ที่ได้จาก HGAs แบบที่ทำวิธี MCS ก่อนแล้วจึงทำ GAs กับ HGAs แบบที่ทำ GAs ก่อนแล้วจึงทำวิธี MCS เพื่อแสดงให้เห็นว่าวิธีที่วิทยานิพนธ์นี้เลือกใช้เป็นวิธีที่มีประสิทธิภาพมากที่สุด

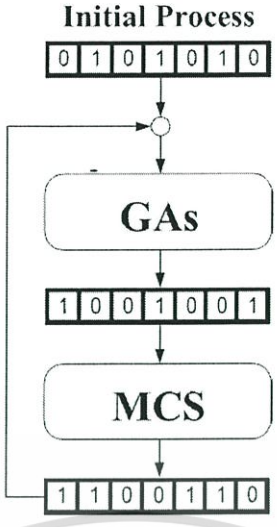
ขั้นตอนการทำงานของ HGAs แบบที่ทำวิธี MCS ก่อนแล้วจึงทำ GAs (โดยต่อไปจะเรียกว่า HGAs แบบที่ ซ.1) จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton", เทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วย โดยสามารถแสดงเป็นแผนภาพการทำงานดังแสดงในรูปที่ ซ.1



รูปที่ ซ.1 แผนภาพการทำงานของ HGAs แบบที่ ซ.1

ส่วนขั้นตอนการทำงานของ HGAs แบบที่ทำ GAs ก่อนแล้วจึงทำวิธี MCS (โดยต่อไปจะเรียกว่า HGAs แบบที่ ซ.2) จะอ้างอิงตามหัวข้อที่ 4.2 ในบทที่ 4 เช่นกัน กล่าวคือจะใช้กระบวนการ MCS แบบที่เลือกโปรโตไทป์จาก Candidate Set เท่านั้น และมีการนำเทคนิค "Singleton", เทคนิค "Diversity Crossover" และเทคนิค "Diversity Selection" เข้ามาทำงานร่วมด้วย แต่จะทำการสลับลำดับการทำงานระหว่างวิธี MCS กับ GAs โดยสามารถแสดงเป็นแผนภาพการทำงานดังแสดงในรูปที่ ซ.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ซ.2 แผนภาพการทำงานของ HGAs แบบที่ ซ.2

โดยผลลัพธ์จะแสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูลทั้ง 6 ชุด โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที, 30 นาที, 1 ชั่วโมง, 3 ชั่วโมง, 6 ชั่วโมง และ 12 ชั่วโมง และเนื่องจาก GAs และ HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละการทำงานจะให้ผลลัพธ์ไม่เหมือนกัน ดังนั้นจึงทำการทดลอง 5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่เจอจะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ซ.1 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ลำดับการทำงานของ HGAs แบบที่ ซ.1

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	-เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:04.90	9.0	00:00:09.18	7.0	00:19:59.85
ECOLI (336)	87.8	01:39:09.46	82.8	03:37:37.01	80.4	03:20:53.29
CREDIT (1000)	457.6	00:01:32.69	451.2	03:48:37.40	435.4	01:40:17.66
YEAST (1484)	869.0	00:07:07.28	869.0	00:07:07.28	853.2	00:05:02.13
MUSK (6598)	792.0	02:33:41.32	761.0	02:29:33.59	749.6	01:58:10.14
PEN (10098)	311.0	01:46:56.29	282.4	01:36:15.39	282.4	01:36:15.39

จากตารางที่ ซ.1 เป็นตารางผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs แบบที่ ซ.1 ของทั้ง 6 ชุดข้อมูล โดยช่องจำนวนโปรโตไทป์แสดงจำนวนโปรโตไทป์ผลลัพธ์ของ HGAs แบบที่ ซ.1 โดยจะพิจารณาจากค่าเฉลี่ยของจำนวนโปรโตไทป์ผลลัพธ์จากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 12 ชั่วโมง) และช่องร้อยละแสดงร้อยละของจำนวนโปรโตไทป์ผลลัพธ์เทียบกับจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ ส่วนตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

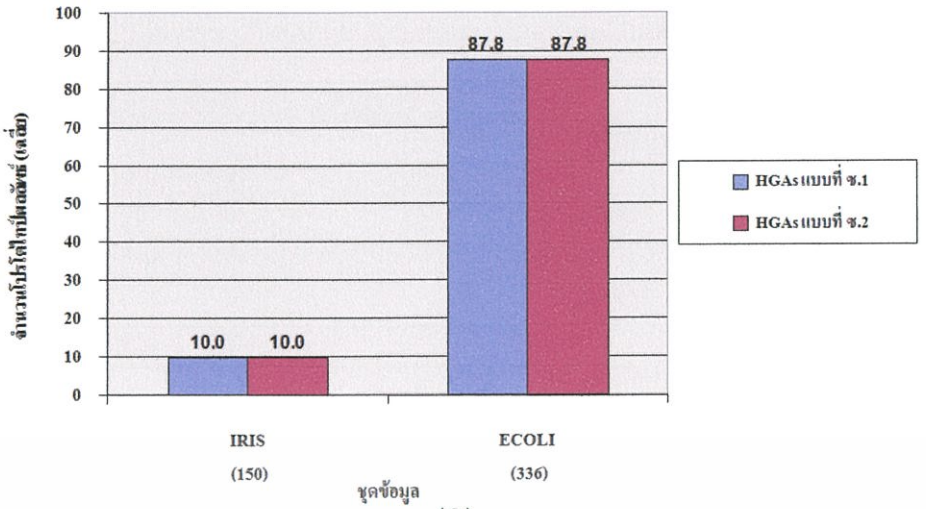
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ซ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ลำดับการทำงานของ HGAs แบบที่ ซ.2

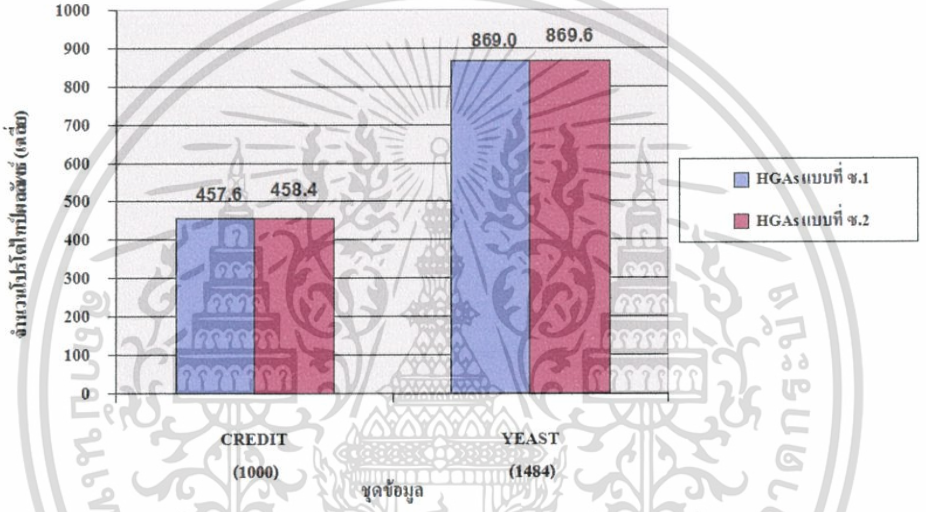
ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
IRIS (150)	10.0	00:00:11.01	9.0	00:00:38.66	7.0	00:07:35.04
ECOLI (336)	87.8	09:24:49.79	83.4	10:11:47.75	80.6	10:17:03.00
CREDIT (1000)	458.4	00:02:53.04	453.0	03:07:24.18	437.2	05:31:21.58
YEAST (1484)	869.6	00:18:59.36	869.6	00:18:59.36	853.2	02:03:32.84
MUSK (6598)	792.4	03:50:51.81	761.2	03:03:33.33	751.0	02:50:00.09
PEN (10098)	319.8	02:04:47.76	269.2	05:19:52.33	256.4	04:53:56.54

จากตารางที่ ซ.2 เป็นตารางผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของ HGAs แบบที่ ซ.2 ของทั้ง 6 ชุดข้อมูล โดยช่องจำนวนโปรโตไทป์แสดงจำนวนโปรโตไทป์ผลลัพธ์ของ HGAs แบบที่ ซ.2 โดยจะพิจารณาจากค่าเฉลี่ยของจำนวนโปรโตไทป์ผลลัพธ์จากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 12 ชั่วโมง) และช่องร้อยละแสดงร้อยละของจำนวนโปรโตไทป์ผลลัพธ์เทียบกับจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ ส่วนตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

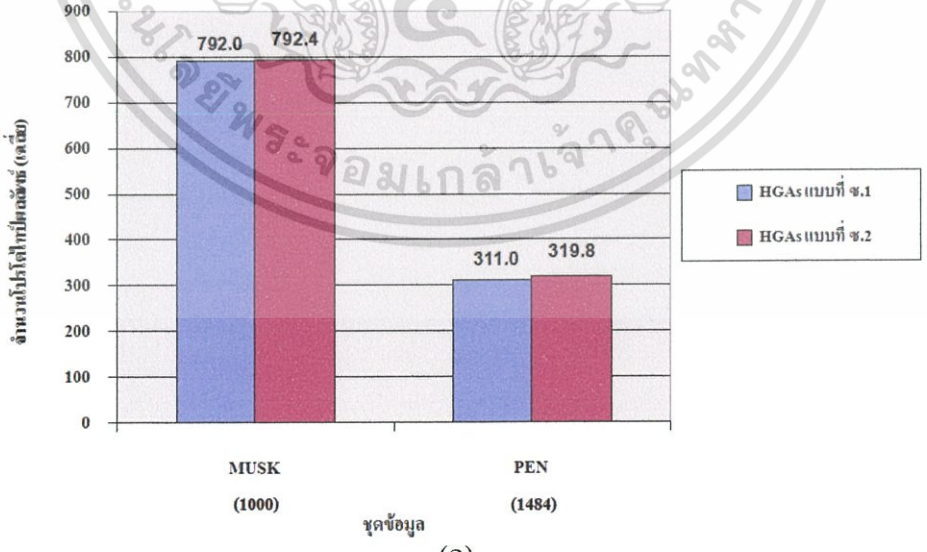
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



(ข)



(ค)

รูปที่ ซ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของลำดับการทำงาน (เฉพาะกรณีสอดคล้อง 100%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ซ.1 เป็นการแสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของลำดับการทำงาน (เฉพาะกรณีสอดคล้อง 100%) โดยรูป (ก) แสดงผลการทดลองชุดข้อมูลขนาดเล็ก ส่วนรูป (ข) แสดงผลการทดลองชุดข้อมูลขนาดกลาง และรูป (ค) แสดงผลการทดลองชุดข้อมูลขนาดใหญ่ โดยแนวแกนตั้งแสดงถึงจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ยที่ได้จากแต่ละวิธี ส่วนแนวแกนนอนแสดงถึงชุดข้อมูลที่ใช้ในการทดลอง และตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

จากตารางที่ ซ.1 และตารางที่ ซ.2 จะเห็นว่าผลลัพธ์ของ HGAs ทั้ง 2 แบบส่วนมากใกล้เคียงกันในทุกชุดข้อมูล ทุกกรณีความสอดคล้อง แต่เวลาการทำงานของ HGAs แบบ ซ.2 จะนานกว่า HGAs แบบ ซ.1 ยกเว้นในชุดข้อมูล PEN ที่ในกรณีความสอดคล้อง 100% นั้น HGAs แบบ ซ.1 ให้ผลลัพธ์ที่ดีกว่า HGAs แบบ ซ.2 แต่ในกรณีความสอดคล้อง 99% และ 98% นั้น HGAs แบบ ซ.2 กลับให้ผลลัพธ์ที่ดีกว่า HGAs แบบ ซ.1

ทั้งนี้อาจจะเป็นเพราะในขั้นตอนการสุ่มกำหนดค่าให้กับประชากรเริ่มต้นของ HGAs ได้มีการกำหนดให้โครโมโซม 1 ตัวมีค่าบิตเป็น “1” ทั้งหมด ซึ่งวิธีการทำงานของวิธี MCS แบบดั้งเดิมที่ใช้ออบเจ็คทั้งหมดในชุดข้อมูลฝึกสอนในการเริ่มต้นการทำงาน ซึ่งแม้ว่าจะทำให้ได้ผลลัพธ์ที่ไม่ใช่ผลลัพธ์ที่ดีที่สุด แต่ก็จะช่วยให้ HGAs ทราบถึงบริเวณที่น่าจะมีคำตอบที่ดีอยู่ ซึ่งอาจจะลู่ไปสู่ผลลัพธ์ที่ดีที่สุดได้ แต่เมื่อทำ GAs ก่อนแล้วจึงทำวิธี MCS อาจจะทำให้โครโมโซมที่มีค่าบิตเป็น “1” ทั้งหมดไม่ผ่านกระบวนการ MCS ทำให้ HGAs ใช้เวลาลู่ไปสู่คำตอบที่ดีนานขึ้น หรืออาจจะไม่ลู่ไปสู่คำตอบที่ดีเลยก็ได้

แม้ว่าผลลัพธ์ของ HGAs แบบที่ ซ.2 ในกรณีความสอดคล้อง 99% และ 98% ของชุดข้อมูล PEN จะดีกว่า HGAs แบบที่ ซ.1 แต่การเลือกโปรโตไทป์ในวิทยานิพนธ์ฉบับนี้จะยึดคุณสมบัติความสอดคล้องเป็นหลัก ดังนั้นวิทยานิพนธ์ฉบับนี้นำเสนอ HGAs แบบที่ทำวิธี MCS ก่อนแล้วจึงทำ GAs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ฌ.

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs

การทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าผลลัพธ์ตัวแปรของ GAs เป็นการทดลองเพื่อเปรียบเทียบผลลัพธ์ที่ได้จาก GAs แบบที่กำหนดค่าตัวแปรต่างกัน เพื่อแสดงให้เห็นถึงผลกระทบของตัวแปรที่มีต่อประสิทธิภาพของการทำงาน

ปัจจัยที่มีผลต่อประสิทธิภาพการทำงานของ GAs มีด้วยกัน 2 ส่วน ส่วนแรก คือ ฟังก์ชั่นคำนวณค่าความเหมาะสมที่จะแตกต่างกันไปขึ้นกับแต่ละปัญหา ส่วนที่สอง คือ การกำหนดค่าตัวแปรต่างๆ โดยตัวแปรพื้นฐานของ GAs มีด้วยกัน 4 ตัวดังนี้

1. จำนวนประชากร (N_{pop})
2. ค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น (P_{ini})
3. ค่าความน่าจะเป็นในการครอสโอเวอร์ (P_c)
4. ค่าความน่าจะเป็นในการกลายพันธุ์ (P_m)

โดยจำนวนประชากร คือ จำนวนโครโมโซมที่ใช้ในการค้นหาคำตอบในแต่ละเจเนอเรชั่น ดังนั้นยิ่งจำนวนประชากรมากก็จะยิ่งมีโครโมโซมที่ใช้ในการค้นหาคำตอบมาก แต่ก็จะต้องเสียเวลาในการทำงานในแต่ละเจเนอเรชั่นเพิ่มขึ้นด้วย ดังนั้นจึงทำการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs ในกรณีที่กำหนดจำนวนประชากรแตกต่างกัน

ส่วนค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้น คือ ค่าความน่าจะเป็นในการกำหนดค่าแต่ละบิตในโครโมโซมให้มีค่าเป็น "1" และเนื่องจากในปัญหาการเลือกโปรโตไทป์ในวิทยานิพนธ์ฉบับนี้ยึดหลักคุณสมบัติความสอดคล้องเป็นหลัก ดังนั้นจึงควรกำหนดค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้นให้มีค่ามากๆ ตามที่ได้อธิบายในบทที่ 4 หัวข้อที่ 4.5.1 ดังนั้นในการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs จึงกำหนดค่าความน่าจะเป็นในการกำหนดค่าประชากรเริ่มต้นเท่ากับ "0.8" เสมอ

ส่วนค่าความน่าจะเป็นในการครอสโอเวอร์ คือ ค่าความน่าจะเป็นในการกำหนดว่าแต่ละบิตในโครโมโซมจะเกิดการครอสโอเวอร์หรือไม่ โดยกระบวนการครอสโอเวอร์ทำหน้าที่เพิ่มจำนวนโครโมโซมที่ใช้ในการค้นหาคำตอบให้มากขึ้น และเนื่องในวิทยานิพนธ์ฉบับนี้ใช้กระบวนการครอสโอเวอร์แบบ "Uniform Crossover" ซึ่งเป็นการสุ่มสลับค่าบิตระหว่างโครโมโซมพ่อ-แม่ ด้วยด้วยอัตราส่วนเท่าๆ กันดังนั้นในการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs จึงกำหนดค่าความน่าจะเป็นในการครอสโอเวอร์เท่ากับ "0.5" เสมอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนค่าความน่าจะเป็นในการกลายพันธุ์ คือ ค่าความน่าจะเป็นในการกำหนดว่าแต่ละบิตในโครโมโซมจะเกิดการกลายพันธุ์หรือไม่ โดยกระบวนการกลายพันธุ์ทำหน้าที่เปลี่ยนแปลงค่าบิต ทำให้โครโมโซมสามารถกระจายการค้นหาคำตอบออกไป ซึ่งเป็นปัจจัยสำคัญที่ช่วยให้โครโมโซมข้าม Local Optimum ไปได้ ซึ่งในวิทยานิพนธ์ฉบับนี้ใช้กระบวนการกลายพันธุ์แบบ "Inverse-Bit Mutation" ซึ่งเป็นการสุ่มกลับค่าบิตจาก "0" เป็น "1" หรือ จาก "1" เป็น "0" ด้วยค่าความน่าจะเป็นในการกลายพันธุ์ ดังนั้นจึงทำการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs ในกรณีที่กำหนดค่าความน่าจะเป็นในการกลายพันธุ์แตกต่างกัน

ขั้นตอนการทำงานของ GAs จะอ้างอิงตามหัวข้อที่ 3.2 ในบทที่ 3 และมีการกำหนดตัวแปรต่างๆ 4 แบบ ดังนี้

ตารางที่ ฌ.1 ค่าตัวแปรของ GAs แบบต่างๆ

GAs แบบที่	N_{pop}	P_{ini}	P_c	P_m
ฌ.1	40	0.8	0.5	0.015
ฌ.2	80	0.8	0.5	0.015
ฌ.3	150	0.8	0.5	0.015
ฌ.4	80	0.8	0.5	0.05

ผลลัพธ์จะแสดงเป็นจำนวนโปรโตไทป์ผลลัพธ์ที่ดีที่สุด (แยกตามกรณีความสอดคล้อง) และเวลาที่ได้ผลลัพธ์ เมื่อทดลองกับชุดข้อมูล YEAST, ชุดข้อมูล MUSK และชุดข้อมูล PEN โดยทำการบันทึกผลลัพธ์ที่ได้เมื่อทำงานไปเป็นเวลา 15 นาที และ 30 นาที และเนื่องจาก HGAs เป็นวิธีการเชิงสุ่มที่ผลลัพธ์ที่ได้ในแต่ละครั้งจะไม่เหมือนกัน ดังนั้นจึงทำการทดลอง5 ครั้ง แล้วพิจารณาผลจากค่าเฉลี่ย

เวลาที่ได้ผลลัพธ์จะแสดงในรูปของ HH:MM:SS.XX โดย HH คือ หน่วยเวลาเป็นชั่วโมง MM คือ หน่วยเวลาเป็นนาที SS คือ หน่วยเวลาเป็นวินาที และ XX คือ หน่วยเวลาเป็นหนึ่งในร้อยของวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ฌ.2 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.1

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
YEAST (1484)	1484.0	00:00:00.00	1106.2	00:24:27.69	1091.8	00:23:43.39
MUSK (6598)	4962.8	00:15:28.80	3760.4	00:23:12.37	3760.4	00:23:12.37
PEN (10092)	4185.2	00:29:19.16	4156.8	00:28:50.76	4156.8	00:28:50.76

ตารางที่ ฌ.3 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.2

ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
YEAST (1484)	1484.0	00:00:00.00	1131.4	00:25:04.91	1115.2	00:19:45.47
MUSK (6598)	5069.2	00:16:39.59	3929.8	00:28:02.60	3929.8	00:28:02.60
PEN (10092)	4042.2	00:29:27.86	4013.4	00:29:03.01	4013.4	00:29:03.01

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ฌ.4 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.3

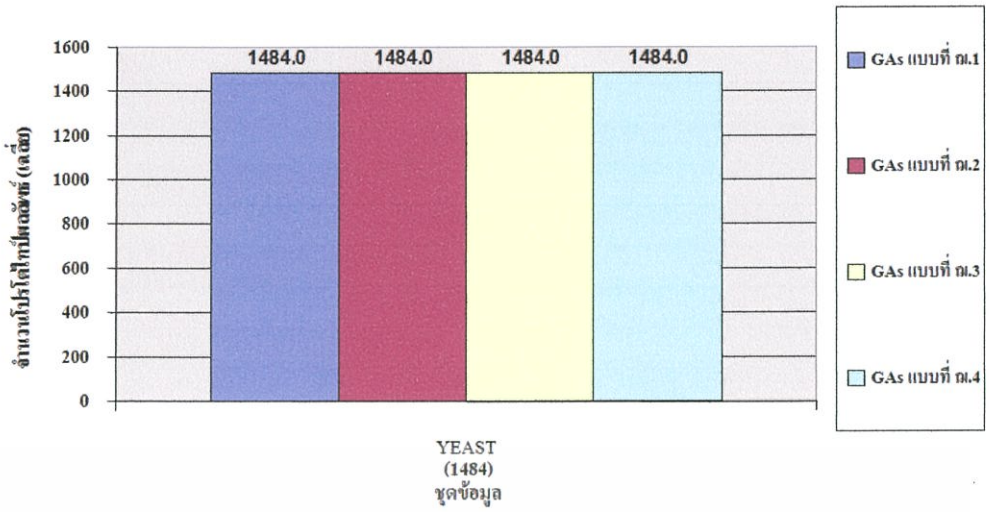
ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
YEAST (1484)	1484.0	00:00:00.00	1144.4	00:27:17.43	1126.4	00:22:26.84
MUSK (6598)	6598.0	00:00:00.00	4029.0	00:28:27.69	4029.0	00:28:27.69
PEN (10092)	4326.2	00:29:37.05	4282.8	00:29:33.06	4282.8	00:29:33.06

ตารางที่ ฌ.5 สรุปผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าตัวแปรของ GAs แบบที่ ฌ.4

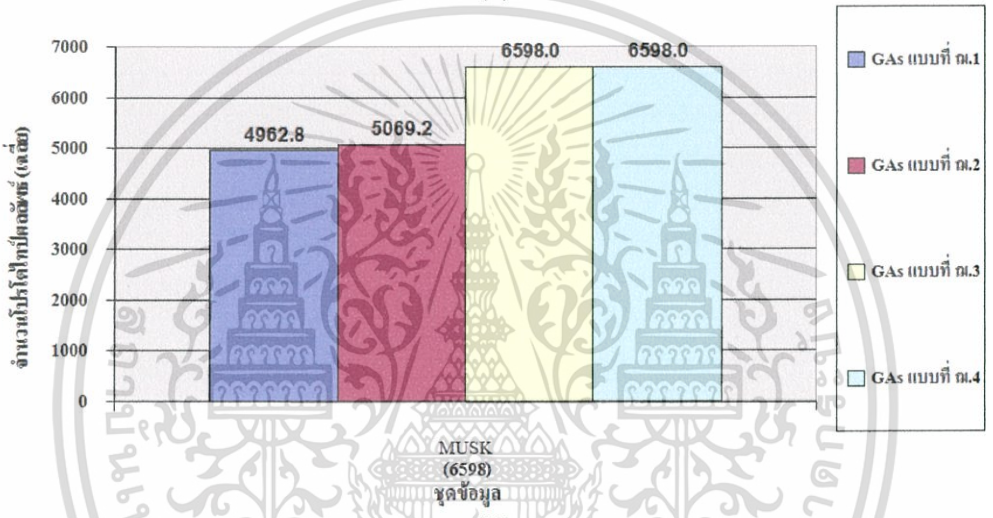
ชุดข้อมูล	สอดคล้อง 100%		สอดคล้อง 99%		สอดคล้อง 98%	
	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)	จำนวนโปรโตไทป์ (เฉลี่ย)	เวลา (เฉลี่ย)
YEAST (1484)	1484.0	00:00:00.00	1484.0	00:00:00.00	1484.0	00:00:00.00
MUSK (6598)	6598.0	00:00:00.00	4032.8	00:27:15.90	4024.0	00:28:11.06
PEN (10092)	5954.4	00:26:28.27	5806.6	00:28:42.81	5806.6	00:28:42.81

ตารางที่ ฌ.1 , ตารางที่ ฌ.2, ตารางที่ ฌ.3 และตารางที่ ฌ.4 เป็นตารางผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าผลลัพธ์ตัวแปรของ GAs ของ GAs แบบที่ ฌ.1, GAs แบบที่ ฌ.2, GAs แบบที่ ฌ.3 และ GAs แบบที่ ฌ.4 ของทั้ง 3 ชุดข้อมูล (ตามลำดับ) โดยช่องจำนวนโปรโตไทป์ และเวลา แสดงจำนวนโปรโตไทป์ผลลัพธ์ และเวลาการทำงาน (ตามลำดับ) โดยทั้งจำนวนโปรโตไทป์และเวลาการทำงานจะพิจารณาจากค่าเฉลี่ยจากการทดลองทั้ง 5 ครั้ง (แต่ละครั้งทำงานเป็นเวลา 30 นาที)

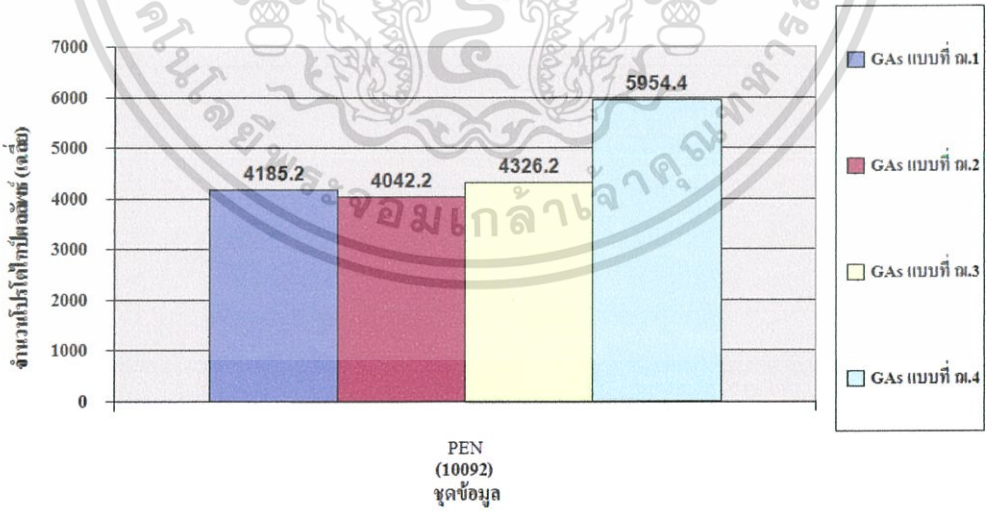
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



(ข)



(ค)

รูปที่ ฌ.1 ผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าผลลัพธ์ตัวแปรของ GAs (เฉพาะกรณีสอดคล้อง 100%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ฌ.1 เป็นการแสดงผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ของค่าผลลัพธ์ตัวแปรของ GAs (เฉพาะกรณีสอดคล้อง 100%) โดยรูป (ก) แสดงผลการทดลองชุดข้อมูล YEAST ส่วนรูป (ข) แสดงผลการทดลองชุดข้อมูล MUSK และรูป (ค) แสดงผลการทดลองชุดข้อมูล PEN โดยแนวแกนตั้งจะแสดงถึงจำนวนโปรโตไทป์ผลลัพธ์เฉลี่ยที่ได้จากแต่ละวิธี ส่วนแนวแกนนอนแสดงถึงชุดข้อมูลที่ใช้ในการทดลอง และตัวเลขในวงเล็บใต้ชื่อชุดข้อมูลแสดงถึงจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้นๆ

จากตารางที่ ฌ.1 ตารางที่ ฌ.2 และตารางที่ ฌ.3 จะเห็นว่าเมื่อทำการเพิ่มจำนวนประชากรจาก “40” เป็น “80” ผลลัพธ์ที่ได้ใกล้เคียงกัน (ดีขึ้นเล็กน้อยในชุดข้อมูล PEN แต่แย่ลงเล็กน้อยในชุดข้อมูล MUSK) แต่เมื่อเพิ่มจำนวนประชากรจาก “80” เป็น “150” ผลลัพธ์ที่ได้กลับแย่ลงอย่างชัดเจนในทุกชุดข้อมูล และเมื่อพิจารณาจากตาราที่ ฌ.4 จะเห็นว่าเมื่อทำการเพิ่มค่าความน่าจะเป็นในการกลายพันธุ์จาก “0.015” เป็น “0.05” (จำนวนประชากรเท่ากับ “80”) ผลลัพธ์ก็แย่ลงอย่างชัดเจน

แม้ว่าผลลัพธ์ในชุดข้อมูล PEN ของ GAs แบบที่ ฌ.2 จะดีกว่า GAs แบบที่ ฌ.1 (ซึ่งเป็น GAs แบบที่ใช้ในการทดลองในบทที่ 5) แต่เมื่อพิจารณาร่วมกับ GAs แบบที่ ฌ.3 ก็จะเห็นว่าการเพิ่มจำนวนประชากร ไม่ได้ทำให้ได้ผลลัพธ์ที่ดีขึ้นเสมอไป จึงสรุปได้ว่าแม้จำนวนประชากรจะเพิ่มขึ้นก็ไม่ได้รับประกับว่าจะทำให้ได้ผลลัพธ์ที่ดีขึ้นเสมอไป และเมื่อพิจารณา GAs แบบที่ ฌ.4 ก็จะเห็นว่าการเพิ่มค่าความน่าจะเป็นในการกลายพันธุ์ก็ไม่ได้ช่วยให้ได้ผลลัพธ์ที่ดีขึ้นเช่นกัน โดยส่วนมากค่าความน่าจะเป็นในการกลายพันธุ์จะถูกกำหนดให้มีค่าน้อยๆ เพราะต้องการให้โครโมโซมไม่พัฒนาเร็วจนก้าวข้ามคำตอบที่ดีไป ดังนั้นการเพิ่มค่าความน่าจะเป็นในการกลายพันธุ์มากเกินไปจึงอาจส่งผลทำให้ไดผลลัพธ์ที่แย่ลง

ดังนั้นจึงสรุปว่าการกำหนดค่าตัวแปรต่างๆ ที่เหมาะสมของ GAs นั้นจะขึ้นกับแต่ละชุดข้อมูล โดยที่ไม่ขึ้นกับจำนวนข้อมูล กล่าวคือแม้ชุดข้อมูลจะมีข้อมูลจำนวนมาก แต่การกำหนดค่าตัวแปรต่างๆ ให้มากขึ้น ก็ไม่ได้ทำให้ได้ผลลัพธ์ที่ดีขึ้น นั่นหมายความว่าไม่มีวิธีการกำหนดค่าตัวแปรต่างๆ ที่เหมาะสมกับทุกชุดข้อมูลได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ญ.

งานวิจัยที่ได้รับการตีพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

http://www.sice.or.jp/sice2008/

SICE Annual Conference 2008

International Conference on Instrumentation, Control and Information Technology

Final Program and Abstracts

Aug.20(Wed.)~22(Fri.)

The University of Electro-Communications(UEC), Chofu, Tokyo, JAPAN

Organized by
The Society of Instrument and Control Engineers (SICE), Japan

Supported by
Chofu City, The Univ. of Electro-Communications, Japan

Technically Co-Sponsored by
IEEE/IES, IEEE/RAS, IEEE/CSS, IEEE/SMC, The Instrumentation, Systems and Automation Society (ISA), Institute of Control, Robotics and Systems(ICROS).

In association with
China Instrument and Control Society (CIS), Chinese Association of Automation (CAA), Chinese Automatic Control Society (CACS), International Measurement Confederation (IMEKO), IEEE Japan Council, IFAC NMO-Japan

SICE UEC IEEE ISA ICROS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SICE Annual Conference 2008
August 20-22, 2008, The University Electro-Communications, Japan

Prototype Selection based on Minimal Consistent Subset and Genetic Algorithms

Boontee Kruatrachue[†] and Marut Hongsamart[*]
Department of Computer Engineering, Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
E-mail: booontee@yahoo.com[†], mnemonic329@hotmail.com[*]

Abstract - **This paper applies the genetic algorithms to identify the minimal "consistent" prototype subset [1]. This subset can be used as a prototype which correctly recognizes the entire original prototype set. This proposed genetic algorithm tries to find the minimal consistent subset to reduce recognition time in nearest neighbor [2] classification. The main difference from other genetic algorithm (GA) approaches is the hybrid of minimal consistent set identification (MCSI) method [3] and genetic algorithm. The MCSI method provides the local optimal number of prototype while the Genetic performs the global search. The proposed hybrid algorithm has been tested on several problems and compared with the results of MCSI and other GA approach [4].**

Keywords – **prototype selection, minimal consistent subset, genetic algorithms, consistency property, nearest neighbor rule.**

I. Introduction

In this paper, we investigate the problem of minimal prototype selection from the training data set with the consistency property [1]. The reduction prototype set is consistence, if the set can be used to correctly classify all the training data using nearest neighbor rule.

There are many researches on this topic which can be divided into "editing" and "condensing" depending on the objective of data removing or "replacement" and "selection" depending on the existence of prototype in original data set.

In this context, we focus on condensing-selection prototype selection with consistency property. This problem is a hard combinatorial problem [6], which is NP problem that can't be solving for exact solution in limited time. Minimal consistent set identification (MCSI) by Dasaraty is one of the prototype selections that tries to obtain minimal consistent set which is usually trapped in a local optimum one. Genetic algorithms (GA's) are also applied for prototype selection. But they are suitable for very small datasets. In order to obtain the benefit of both methods, this paper proposed how to combine them to utilize the global search of GA's and local search of MCSI.

This paper is organized into 5 sections. The nearest neighbor classification is reviewed in section II. Section III and IV review minimal consistent set identification and genetic algorithms for prototype selection, respectively. Section V, explains the proposed hybrid of MCSI and genetic algorithm. The experimental results are shown in section VI. Section VII is the conclusions.

II. Nearest Neighbor Rule

A prototype set $((x_1,y_1) , c_1) , \dots , ((x_n,y_n) , c_n)$ is given, where the (x_i,y_i) is the attribute value of the data point i^{th} and c_i is the category of the data i^{th}.

The unknown data ((a,b) , z) can be classified as c_{nn}, which is the category of the prototype that is the nearest neighbor of (a,b). If d((x,y) , (a,b)) is the distance function that measure the difference (x,y) and (a,b). The nearest neighbor of unknown data at (a,b) is defined as

$$(x_{nn},y_{nn}) = \min d((x_i,y_i) , (a,b)) ; i = 1, 2, \dots , n \quad (1)$$

III. Minimal Consistent Set Identification

Minimal consistent set identification (MCSI) is one of the condensing-selection prototype selections, which based on the concept of covering defined by "NUN", the Nearest Unlike Neighbor [3]. Any data point A is covered by any data point B which has the same class as A, as long as B is closer to A than A's NUN. Hence, A can be correctly classified, using nearest neighbor rule, to be the same class as B if B is selected as a prototype, as B is closer to A than A's NUN.

Once a distance matrix among all the data in the training set is sorted and all NUNs of each data point are located, a list of data points covered by any point B, B cover set, can be constructed. The cover list of B includes any of the data point A that has B located closer to A than A's NUN. The prototypes can be greedy selected incrementally by choosing a data point which has the largest cover list. Once that prototype is selected all the data point in that data cover list is covered (guarantee to be correctly classified) and that cover list is subtracted from all the cover lists. Then the data with the largest cover list is selected again as another prototype. The selection and subtraction process continues until all the cover lists are empty, which indicate the occurring of consistence property, where all the training data can be correctly classified by the selected prototypes.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The prototype selection from the above algorithm is still not minimal consistency due to two reasons. The first one is the greedy order of prototype selection which has no guarantee to minimal number of prototype. The second one is the inaccuracy of cover list calculation using NUN as boundary, since if the NUN is not selected as a prototype the cover lists can be expand. But before all prototypes are selected, it is safe to assume cover list boundary at NUN. To alleviate this problem, the selected prototypes are used again as boundary (instead of NUN) and the whole process repeated until no change in prototype set.

The MCSI method proceeds in the following steps as shown in Fig. 1.

1) Create a sorted distance table (c) among all data in the original prototype set (a).

2) Label all the NUN (Nearest Unlike Neighbor) by assumes that all the NUN are selected prototype (c).

3) Create a cover list for each prototype (d).

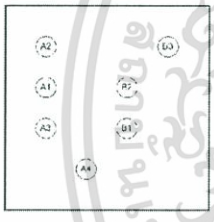
4) Select the prototype with maximum cover, and remove all the data in the covered list from all the cover lists (e).

5) Repeat step 4 until all the cover lists is empty and obtains all new selected prototypes (f)

6) Stop if the new selected prototypes from step 5 are the same as the previous one.

7) Re-Label the sorted distance table (g) with the new selected prototypes from step 5.

8) Go to step 3.



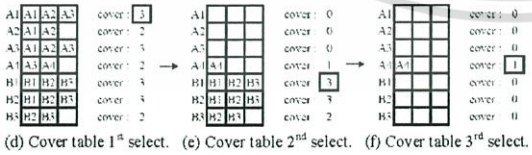
(a) Original prototype set.

	A1	A2	A3	A4	B1	B2	B3
A1	0.0	1.0	1.0	2.2	2.2	2	3.1
A2	1.0	0.0	2.0	3.1	2.8	2.2	3.0
A3	1.0	2.0	0.0	1.4	2.0	2.2	3.6
A4	2.2	3.1	1.4	0.0	1.4	2.2	3.5
B1	2.2	2.8	2.0	1.4	0.0	1	2.2
B2	2	2.2	2.2	2.2	1	0.0	1.4
B3	3.1	3.0	3.6	3.6	2.2	1.4	0.0

(b) Distance table.

A1	A1	A2	A3	B2	B1	A4	B3
A2	A2	A1	A3	B2	B1	B3	A4
A3	A3	A1	A4	B1	A2	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

(c) 1st Sorted and labeled distance table.



(d) Cover table 1st select. (e) Cover table 2nd select. (f) Cover table 3rd select.

A1	A1	A2	A3	B2	B1	A4	B3
A2	A2	A1	A3	B2	B1	B3	A4
A3	A3	A1	A4	B1	A2	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

(g) Re-label sorted distance table.

Fig. 1. Illustration of MCSI method.

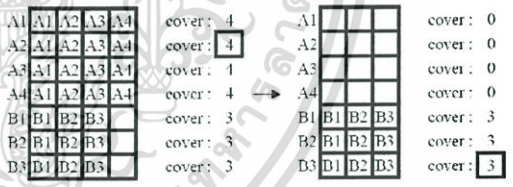
The "cover" concept is shown in Fig. 1(c) and 1(d), where A1 cover A1, A2, A3 since A1 is closer to A1, A2 and A3 than B2, B2 and B1 (the nearest unlike neighbor of A1,A2 and A3). Hence, if A1 is selected as a prototype, A1, A2 and A3 can be correctly recognized.

From Fig. 1, the first round MCSI obtains A1, A4 and B1 as prototypes. After the second round MCSI obtains the same prototype. So MCSI obtains 3 prototypes (A1, A4 and B1) from 7 data in original dataset.

If assume that only A2 and B3 are selected prototype and used as boundary instead of NUNs as shown in Fig. 2(a). In this case, A2 and B3 are selected as prototypes as shown in Fig. 2(b) and 2(c). Hence, the optimality depends on how to initially label boundary.

A1	A1	A2	A3	B2	B1	A3	B3
A2	A2	A1	A3	B2	B1	B3	A3
A3	A3	A1	A4	B1	A1	B2	B3
A4	A4	B1	A3	B2	A1	A2	B3
B1	B1	B2	A4	A3	A1	B3	A2
B2	B2	B1	B3	A1	A4	A3	A2
B3	B3	B2	B1	A2	A1	A4	A3

(a) Sorted and labeled distance table.



(b) Cover table 1st selection. (c) Cover table 2nd selection.

Fig. 2. Illustration of another NUNS approach.

IV. Genetic Algorithms in Prototype Selection

GA's are evolutionary optimization technique [4]. They are basically random search with some genetic process to guide the search. GA usually perform effective search with the right chromosome representation and proper fitness function.

In minimal consistency set (MCS) problem, fitness function should reward correct recognition accuracy and penalize large prototype set, as shown below.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\mathbb{F}(S) = \frac{ClassifiedNum(S) - \alpha * Size(S)}{N_OriginalDataset} \quad (1)$$

Where ClassifiedNum(S) is classification accuracy of original dataset when using S as the prototype set, Size(S) is number of prototype of S and α is weight between accuracy and size. In this paper we use

$$\alpha = 1/N_OriginalDataset \quad (2)$$

The solution to MCS problem, prototype set S, needs to be represented in the form of "chromosome". In this case, it is encoded into a binary string. Position and length of chromosome are corresponding to the original training dataset, where "1" indicates as selected prototype and "0" as non-selected.

The GA's proceed in the following steps.

1) Initial Process

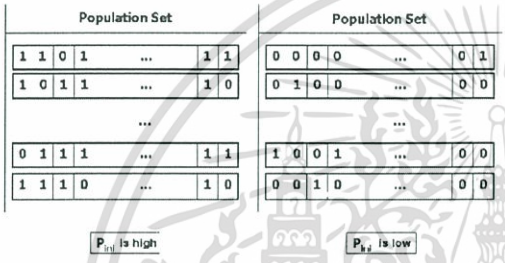


Fig. 3. Illustration of initial process.

A set of chromosomes is randomly generated to be the initial "population set". Each bit in the chromosome takes the value "1" with a prespecified probability P_{ini}. By this parameter, we can generate sparse or dense chromosome for initializing the search.

2) Crossover Process

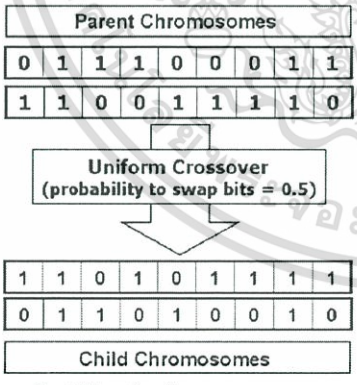


Fig. 4. Illustration of crossover process.

Here we adopted uniform crossover, so the parent chromosomes randomly swap their i^{th} bits with a certain probability "0.5".

3) Mutation Process

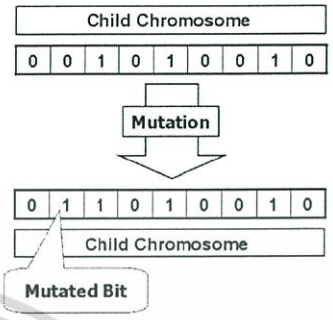


Fig. 5. Illustration of mutation process.

Each bit of each child chromosome changes (mutates) with a prespecified probability (P_m). Here we adopted inverse-bit mutation, so the child chromosomes randomly alternate ("0" will become "1" and "1" will become "0").

4) Selection Process

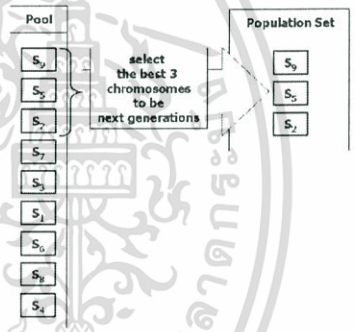


Fig. 6. Illustration of selection process.

All parent and child chromosomes are pooled and the best N_{pop} individuals survive (according to their fitness values). They will be the new population set for the next generation (elitist strategy).

Since GA is based on random procedure, it sometime works on the small-size datasets. In large-size datasets, it tends to obtain prototype set that never has 100% recognition rate and usually has a large set size. There is because GA has no guidance of cover list of each prototype. It is easy to randomly include two prototypes that cover the same data and end up in large prototype set. Also, it is hard to random prototype that has 100% recognition rate since they don't know which prototype cover the uncover data and hardly random the needed one.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

V. HYBRID GENETIC ALGORITHM

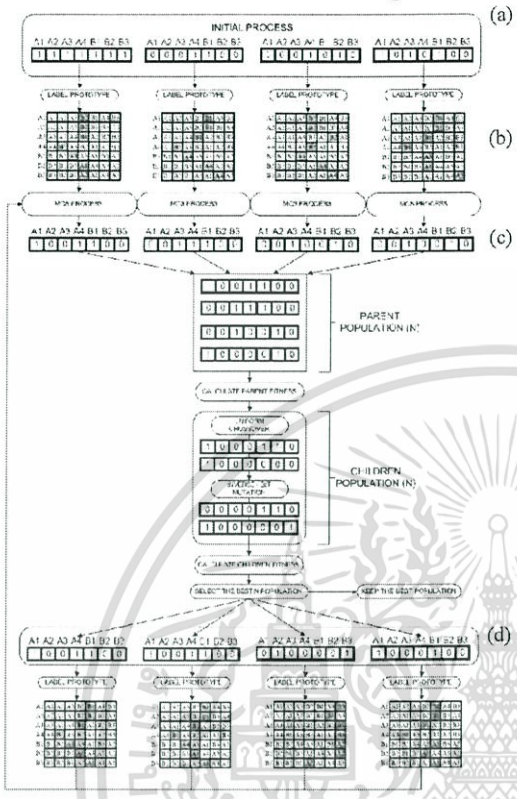


Fig. 7. Illustration of hybrid genetic algorithm.

This paper proposes hybrid genetic algorithm (HGA). This algorithm lessen MCSI's local optimal problem by using GA's to global search. The search performs by random initial selected prototype to set boundary to identify covers for each data in MCSI process.

The problem with MCSI is that the algorithm aim to find a prototype, but from the start, it needs to find cover list which is not possible until the prototype is known. Hence, it initially assumes all prototypes are possible by using all NUNs which cause local optimal. The using of NUNs also causes the MCSI to obtain the same prototype set every run. By varying the starting prototype, MCSI generate different prototype sets each run. The generated prototype set from MCSI usually improves with better recognition rate and smaller set size. Hence, MCSI can be used to improve prototype set generate from GA crossover and mutation process. So, GA can perform the global search while MCSI does the local search.

Hybrid genetic algorithm (HGA) is about the same as normal GA in section IV except the use of single pass MCSI. The MCSI uses prototype set from GA to set boundary in step 2 of MCSI in section II. The distance table among all training data is the same for all population and needs to be calculated only once. Normal MCSI is performed by the identification of cover list for each data and repeatedly greedy selects prototype and subtracts its cover list from all cover lists. The difference with normal MCSI is that it only runs for single pass and always stops at step 6.

The details combination of MCSI and GA is shown in Fig. 7. The first Process is initial process. After random the chromosomes that represent selected prototype, each chromosome is apply to MCSI by setting the boundary for MCSI process as shown in Fig. 7(a) and 7(b). The result prototype sets from MCSI are encoded back to chromosomes and become the population set for GA as shown in Fig. 7(c). The crossover, mutation process are the same as shown in section IV. The winning chromosomes of selection process become the chromosomes for the next generation. These chromosomes are again applied to MCSI process as shown in Fig. 7(d).

In addition, there are 3 more techniques that have been added into HGA to improve its results.

1) MCSI history

Each chromosome also contains the results after applied to MCSI. If the result is the same as the initial chromosome, that chromosome can bypass the MCSI process. This usually happen when the chromosome goes through MCSI process for some times and the local optimum for that chromosome is reached. This saves a lot of run time of MCSI process.

2) Diversity crossover

Many similar chromosomes after pass through MCSI process tend to converge to the same local optimal. In order to safe run time and allow GA to perform wider global search, the "diversity" concept is applied to crossover process. Each parent chromosomes selection by roulette wheel is compared and count number of different bits in their chromosome. They can perform crossover process if their different is greater than some threshold. The threshold is calculated from the average bit different among all chromosomes.

3) Diversity selection

Beside crossover process, "diversity" is also applied in selection process. Instead of select the best chromosomes based on only fitness value, only half of population are selected based on their fitness value and another half based on their diversity to the rest of chromosomes.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

VI. EXPERIMAENTAL RESULTS

The size of prototype set from the proposed hybrid genetic algorithm (HGA) is compared to GA [4] and MCSI [3] on datasets available at UCI site [7]. The test datasets are divided into small and large size datasets. Small datasets consist of IRIS, GLASS and ECOLI .Large datasets consist of PIMA TIC-TAC-TOE and YEAST as shown in Table I.

TABLE I
DETAIL OF DATASETS

	NUMBER OF DATA	DIMENSION	CLASS
IRIS	150	4	3
GLASS	214	9	7
ECOLI	336	7	8
PIMA	768	8	2
TIC-TAC-TOE	956	9	2
YEAST	1484	8	10

TABLE II
PARAMETERS IN GA AND HGA

	N_{pop}	P_{ini}	P_m
GA	40	0.8	0.015
HGA	10	0.8	0.015

Due to random nature of GA, GA and HGA are performed 5 times and use average result. Both algorithms perform on the same amount of time for 15 minutes with parameters as shown in Table I. Where N_{pop} is number of population, P_{ini} is initial probability and P_m is mutation probability.

TABLE III
SMALL DATASET RESULTS OF MCSI, GA AND HGA

	IRIS (150)	GLASS (214)	ECOLI (336)
MCSI	15	85	100
GA	11.6	79.8	95.6
HGA	10.0	79.6	89.2

For small dataset, both GA and HGA outperform MCSI especially in a very small dataset IRIS. In case of IRIS dataset, the minimal consistent set was determined with size 10 by integer nonlinear programming method [8]. Only HGA get the optimum number for all 5 runs. As the data size get larger, the performance different is smaller. In case of ECOLI, HGA is quite better than both GA and MCSI.

TABLE IV
LARGE DATASET RESULTS OF MCS, GA AND HGA

	PIMA (768)	TIC-TAC-TOE (956)	YEAST (1484)
MCSI	325	64	875
GA	330.8	94.0	1484.0
HGA	322.8	63.4	871.0

For larger datasets, GA has the worst performance due to its randomness. Especially in YEAST dataset, GA can not find any consistence set except the original dataset. With some guidance of MCSI, HGA still obtain a little smaller prototype set than MCSI.

VII. CONCLUSIONS

Prototype selection is crucial in order to reduce recognition time in nearest neighbor rule for large-training dataset. MCSI and GA have been used to obtain reduction in prototype set. This paper proposes hybrid genetic algorithm (HGA) which combine MCSI and GA in order to obtain smaller number of prototypes. The HGA take advantage of local optimum of MCSI and wider global search of GA. The test results show a lot of improvements in both small datasets over MCSI and large data set over GA.

REFERENCES

[1] G P.E. Hart, "The Condensed Nearest Neighbor Rule", IEEE Transactions on Information Theory, Vol. 14, 1968, pp. 515-516.

[2] T.M. Cover and P.E. Hart, "Nearest Neighbor Pattern Classification ", IEEE Transactions on Information Theory, Vol. IT-13, No. 1, January 1967, pp. 21-27.

[3] B.V. Dasarathy, "Minimal Consistent Set (MCS) Identification for Optimal Nearest Neighbor Decision Systems Design", IEEE Transactions on Systems, Man and Cybernetics, Vol. 24, No. 3, 1994, pp. 511-517.

[4] L.I. Kuncheva and J.C. Bezdek, "Nearest Prototype Classification: Clustering, Genetic Algorithms, or Random Search?", IEEE Transactions on Systems, Man and Cybernetics, Vol. 28, No. 1, 1998, pp. 160-164.

[5] R. A. Mollineda, F. J. Ferri, and E. Vidal. "Merged-based prototype selection for nearest neighbor classification", Proceedings of the 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI2000), Orlando, USA, July 2000, pp. 640-645.

[6] V. Cerveron and F.J. Ferri, "Another move toward the minimum consistent subset: a tabu search approach to the condensed nearest neighbor rule", IEEE Transactions on Systems, Man and Cybernetics, Vol. 31, No. 3, 2001, pp. 408-413.

[7] Department of Information and Computer Science, University of California, "UCI Machine Learning Repository" [Online]. Available: http://www.ics.uci.edu/~mlearn/ MLRepository.html. 1998.

[8] K. Kangkan, B. Kruatrachue, "Minimal Consistent Subset Selection as Interger Nonlinear Programming Problem", ISCIT 2006, October 2006.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ - นามสกุล	นายมารุต หงษ์สามารถ
วัน เดือน ปีเกิด	19 ตุลาคม 2526
การศึกษา	ปีการศึกษา 2538 – 2543 ระดับมัธยมศึกษา
	โรงเรียนพูลเจริญวิทยาคม
	ปีการศึกษา 2544 - 2547 ระดับอุดมศึกษา
	คณะวิศวกรรมศาสตร์ สาขาวิศวกรรมไฟฟ้า - คอมพิวเตอร์
	มหาวิทยาลัยเทคโนโลยีมหานคร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้