

บทที่ 2

วิวิพันธุกรรมศาสตร์

2.1 บทนำ

วิวิพันธุกรรมศาสตร์เป็นการกระบวนค้้นหาบนพื้นฐานของวิวิฒนาการโดยธรรมชาติ ซึ่งนำเสนอโดย Holland [1] ในปีค.ศ. 1975 ความสามารถกรณีทั่ว ๆ ไปเป็นกระบวนการที่มุ่งหมายวิธีที่ใ้ผลที่ดีที่สุด โดยวิวิพันธุกรรมศาสตร์มีลักษณะเป็นการคัดเลือกในธรรมชาติโดยผู้ที่แข็งแรงกว่าจะเป็นผู้ที่อยู่รอด โดยจะได้สิ่งที่ดีที่สุดที่เกิดขึ้นในรอบของวัฏจักรการใ้กำเนิด ซึ่งในแต่ละรอบของวัฏจักรการใ้กำเนิดนั้นอาจมีวิธีการแก้ปัญหาค้ที่สุดต่างกัน ภายในกระบวนการทางพันธุกรรมศาสตร์นั้นประกอบด้วยสิ่งที่ทำหน้าที่ควบคุมและถ่ายถอดลักษณะเด่นและด้อยของสิ่งมีชีวิตที่เรียกว่า “โครโมโซม” โดยหนึ่งคู่ของโครโมโซมในนิวเคลียสจะเป็นโครโมโซมที่ใ้รับจากฝ่ายพ่อและฝ่ายแม่ที่เป็นผู้ใ้กำเนิด ซึ่งการนำวิธีทางพันธุกรรม ไปใ้แก้ปัญหาค้ในงานต่าง ๆ นั้นจะประกอบไปด้วยกลุ่มของพารามิเตอร์ต่าง ๆ ที่เปรียบเสมือนโครโมโซม และพารามิเตอร์แต่ละตัวเปรียบเสมือนยีน ใ้มีการนำวิธีนี้ไปใ้ประโยชน์มากขึ้นในงานด้านธุรกิจ วิทยาศาสตร์ และวิศวกรรมศาสตร์ [2] เพราะเป็นวิธีที่มีประโยชน์ในค้้นหาการปัญหาค้ที่มีทางออกหลาย ๆ ทาง และยังเป็นวิธีที่ค้ดแปลงและแก้ใ้ง่ายในการนำไปใ้ในงานต่าง ๆ

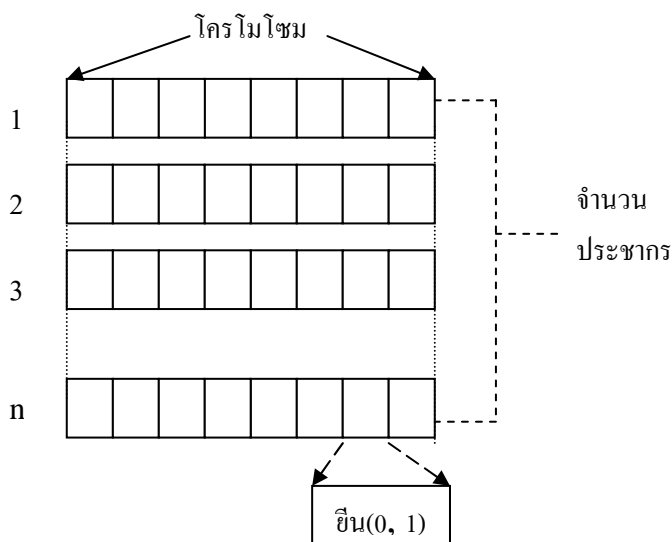
วิวิพันธุกรรมศาสตร์โดยหลัก ๆ จะเป็นการค้้นหาทางออกของปัญหาค้ในรอบแห่งการใ้กำเนิด (Generations) ในแต่ละรอบ ซึ่งองค์ประกอบในวิวิพันธุกรรมศาสตร์มี การสร้างประชากรเริ่มแรก (Initialization) การหาค้ค่า (Evaluation) การคัดเลือก (Selection) ตัวดำเนินการพันธุกรรม (Genetic Operator) ในขั้นตอนแรกเป็นการสร้างประชากรเริ่มต้น เพื่อเป็นตัวแทนแสดงการทางแก้ปัญหาค้ที่มองออกมาในรูปของโครโมโซม เมื่อได้แล้วจะทำการหาค้ค่าของโครโมโซมที่ค้หนดโดยฟังก์ชันฟิตเนส (Fitness Function) ซึ่งมีความเกี่ยวข้องกับฟังก์ชันวัตถุประสงค์ (Objective Function) หรือฟังก์ชันเป้าหมาย โดยการคัดเลือกจะเลือกโครโมโซมที่มีฟังก์ชันวัตถุประสงค์ที่ดีสำหรับการแก้ปัญหาค้ นั้น แล้วใ้ตัวดำเนินการพันธุกรรมศาสตร์เป็นตัวถ่ายเทในการผสม (Recombination) การกลายพันธุ์ (Mutation) และการแทนที่ (Reinsertion) ในรอบของการใ้กำเนิดจนกระทั่งได้ทางแก้ปัญหาค้ที่ค้ถึงในปัญหาค้

2.2 วิวิพันธุกรรมศาสตร์

หลักการของวิวิพันธุกรรมศาสตร์อยู่ในรูปโปรแกรมที่ใ้ใช้งานในการหาทางแก้ปัญหาค้แบบวิธีที่เจาะจง หรือค้ดแปลงใ้เหมาะสมกับสภาพสิ่งแวดล้อมที่ค้ต้องการได้ โดยปัจจัยที่เป็นหลัก

สำคัญนั้นมาจาก Darwinian ซึ่งเป็นการค่อย ๆ วัฒนาการปรับปรุงและคัดเลือกประชากรให้ที่เหมาะสมนำมาเป็นต้นแบบในแต่ละการกำเนิด โดยวิธีพันธุกรรมศาสตร์มีการกำหนดกลุ่มขนาดของประชากรที่ได้รับเลือกในการแก้ปัญหา ซึ่งประชากรที่ได้รับเลือกจะมีการทำการแข่งขันซึ่งกันและกันเพื่อให้อยู่รอดครั้งแล้วครั้งเล่าจนทำให้ได้ผลของการแก้ปัญหาที่ดีตามวัตถุประสงค์ของปัญหาซึ่งขึ้นอยู่กับควบคุมตัวดำเนินการพันธุกรรมศาสตร์ ถ้าวิเคราะห์ตามหลักเหตุผลแล้ววิธีพันธุกรรมศาสตร์เป็นการสะสมทางแก้ปัญหาตลอดเวลาที่มีการค้นหาทางที่ดีกว่ามีผลประโยชน์ที่จะทำให้เข้าสู่คุณภาพที่ต้องการ [3]

Holland เป็นคนแรกที่น่าโปรแกรมวิธีพันธุกรรมศาสตร์มาจำลองในลักษณะกระบวนการประเมินค่าแบบธรรมชาติทางลักษณะชีววิทยาที่ใช้หลักเหตุผลในการหาวิธีแก้ปัญหาที่ดีที่สุด ดังแสดงในรูปที่ 2.1 ซึ่งแสดงโครงสร้างโครโมโซม



รูปที่ 2.1 โครงสร้างโครโมโซม

โครโมโซม (Chromosome) เป็นกลุ่มของยีนหรือจำนวนไบนารีซึ่งแสดงถึงองค์ประกอบของการแก้ปัญหาบนขอบเขตของปัญหานั้น โดยวิธีแก้ปัญหาที่ได้รับการคัดเลือกซึ่งจะอยู่ในกลุ่มของยีน หรือจำนวนไบนารีที่มีลักษณะสำคัญอยู่ที่โครโมโซมในสตริง (String) ไบนารีบิตสามารถเห็นเป็นภาพโครงสร้างของโครโมโซม และมีค่าเป็น 0 และ 1 ที่มีความสัมพันธ์อยู่ในยีนโดยเป็นตัวแสดงลักษณะเฉพาะภายในโครโมโซมแต่ละโครโมโซม

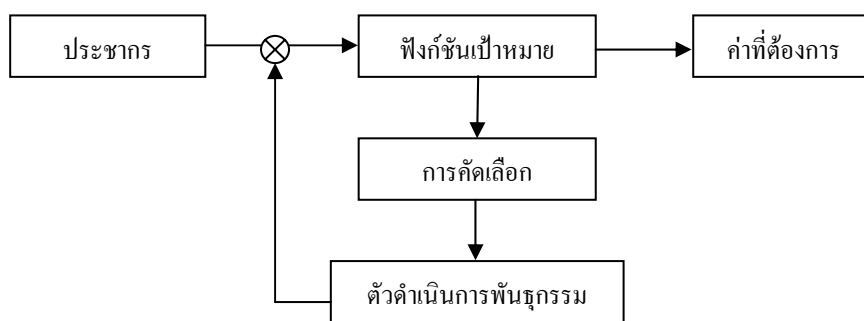
ประชากร (Population) ในการเลือกกลุ่มของโครโมโซมจะเป็นการเลือกจากกองกลาง (Pool) ที่มีขอบเขตภายใต้เงื่อนไขของปัญหา

การให้กำเนิด (Generation) จะเป็นการให้กำเนิดประชากรในแต่ละส่วน ซึ่งขนาดของประชากรจะคงที่สำหรับการให้กำเนิด และแต่ละการให้กำเนิดจะแสดงผลกระทบต่อการปฏิบัติการของวิธีพันธุกรรมศาสตร์

ฟีโนไทป์ (Phenotype) เป็นตัวแสดงรูปแบบคุณสมบัติลักษณะภายนอกที่ปรากฏของสิ่งมีชีวิตโดยกำเนิด ซึ่งได้รับการวิวัฒนาการและสะสมคุณสมบัติภายใต้สิ่งแวดล้อมจนมีการแสดงออกถึงลักษณะภายนอกหรือฟีโนไทป์

ค่าฟิตเนส (Fitness Value) เป็นการหาค่าที่อยู่ในเงื่อนไขของฟังก์ชันฟิตเนส และค่าที่ได้จะเป็นค่าฟิตเนสของการแก้ปัญหาานั้น ๆ

วิวัฒนาการศาสตร์ประกอบด้วยกลุ่มของประชากรแต่ละราย (Individual) และเป็นกลุ่มทางชีววิทยาที่เป็นตัวก่อให้เกิดการกำหนดลักษณะบนประชากร Holland ได้นำวิวัฒนาการศาสตร์เข้ามาที่จะแก้ปัญหาในรูปแบบของสตริงไบนารีบิตของ 0, 1 ในแต่ละสตริงจะแสดงความสามารถที่ซ่อนเร้นของการแก้ปัญหาไว้ ซึ่งในธรรมชาติแล้วจะประกอบด้วยประชากรที่แข็งแรงที่มักจะสืบดำรงชีวิตต่อไปได้ ตลอดจนจะให้กำเนิดประชากรลูกหลาน (Offsprings) ด้วยเหตุนี้จึงมีการถ่ายทอดคุณลักษณะทางชีววิทยาทางสายพันธุ์มายังประชากรลูกหลานที่ให้กำเนิดใหม่นี้ด้วย โดยแต่ละการแก้ปัญหาจะมีความสัมพันธ์อยู่ที่ค่าฟิตเนสที่เป็นส่วนในการวิเคราะห์ว่าประชากรแต่ละรายคืออะไร เมื่อนำไปเปรียบเทียบกับปัญหาอื่นกับประชากรแต่ละรายอื่น ๆ ทำให้มีโอกาสได้ประชากรที่มีคุณลักษณะเด่นที่ดีมา และค่าฟิตเนสที่มากของประชากรแต่ละราย ซึ่งจะทำให้มีโอกาสที่จะมีการถ่ายทอดคุณสมบัติระหว่างประชากรที่มีคุณสมบัติเด่นด้วยกันมาผสมกันในการให้กำเนิดประชากรรุ่นลูกหลานต่อไป โดยจะช่วยส่งเสริมให้ประชากรรุ่นใหม่ให้มีคุณลักษณะเด่นที่ดีขึ้น ดังรูปที่ 2.2 จะเห็นตัวอย่าง โครงสร้างกระบวนการพันธุกรรมศาสตร์ในแต่ละขั้นตอนดังนี้



รูปที่ 2.2 กระบวนการวิวัฒนาการศาสตร์

2.3 ประชากรเริ่มต้นและการแทนที่ประชากร (Initialization and Population Representation)

ในกระบวนการพันธุกรรมศาสตร์จะเริ่มจากการสร้างประชากรเริ่มต้นโดยการสุ่มหรือการเพาะพันธุ์ (Seeded) [4] ซึ่งมีการเข้ารหัสของพารามิเตอร์ในการแก้ปัญหาที่อยู่ภายในประชากรหรือโครโมโซม เพื่อเสนอวิธีแก้ปัญหาให้กับปัญหานั้น โดยในการสร้างจำนวนของประชากรที่จะ

สร้างนั้นเป็นการยากที่จะกำหนดได้ขึ้นอยู่กับงานที่จะใช้ ซึ่งขนาดของจำนวนประชากรมีผลต่อความสามารถในการแก้ปัญหา ถ้าสร้างจำนวนประชากรที่มากย่อมจะได้วิธีการในการแก้ปัญหาที่มากขึ้นและดีขึ้นด้วย แต่จะทำให้ต้องใช้ทรัพยากรในการคำนวณที่มากขึ้น หรือถ้าขอบเขตของการแก้ปัญหาที่ต้องการ เป็นขอบเขตที่มีค่าการเปลี่ยนแปลงที่กว้างจะส่งผลให้ขั้นตอนของงานนั้น ซ้ำออกไปได้ และส่วนที่สำคัญในการพัฒนาการในประชากรของวิวิพันธุ์กรรมศาสตร์ คือการแทนที่ประชากร หรือ โครโมโซม เป็นการเข้ารหัสแบบไบนารี (Binary Code) หรือเลขจำนวนเต็มของแต่ละยีน (Gene) [3] และจะมีการเชื่อมตัวแปรแต่ละตัวเข้าด้วยกันเป็นสตริงไบนารีหรือโครโมโซม โดยส่วนประกอบของยีนจะมีการผสมกันจากตัวดำเนินการพันธุกรรมศาสตร์ ซึ่งต่อมาได้มีการนำการเข้ารหัสแบบเกรย์ (Gray Code) ที่ให้ผลที่ดีกว่าการเข้ารหัสแบบไบนารีมาใช้ และการเข้ารหัสแบบค่าจริง (Real-Value) นี้เสนอโดย Wright [5] โดยการเข้ารหัสแบบค่าจริงจะมีข้อดีที่ไม่ต้องทำการแปลงค่าจากโครโมโซมเป็นฟิโนไทป์ จากการเข้ารหัสในวิธีต่าง ๆ จะช่วยเพิ่มประสิทธิภาพในการทำงานของวิวิพันธุ์กรรมศาสตร์ หรือขึ้นอยู่กับลักษณะของปัญหาและการใช้งาน

2.4 ฟังก์ชันเป้าหมายและฟังก์ชันฟิตเนส (The Objective and Fitness Function)

ฟังก์ชันเป้าหมายถูกกำหนดตามขอบเขตและปัญหาที่จะแก้ไข โดยผู้ใช้ ซึ่งเป็นคุณลักษณะเฉพาะของประชากรแต่ละตัว ในการหาทางแก้ปัญหาที่มีค่าต่ำสุดหรือค่าสูงสุดของปัญหานั้นด้วยการอาศัยฟังก์ชันฟิตเนสที่เป็นตัววัดสถานะของประชากรแต่ละตัว และต้องมีค่าเป็นบวกเสมอ แต่ในบางครั้งฟังก์ชันเป้าหมายอาจจะมีค่าเป็นลบจึงต้องทำการเปลี่ยนดังสมการที่ (2.1)

$$F(x) = g(f(x)) \quad (2.1)$$

โดย $F(x)$ คือ ค่าฟิตเนสที่ได้ g คือ การเปลี่ยนแปลงค่าของฟังก์ชันเป้าหมายเพื่อไม่ให้เป็นลบ และ $f(x)$ คือ ค่าฟังก์ชันเป้าหมาย ในกรณีทั่ว ๆ ไปค่าฟังก์ชันฟิตเนสจะมีความสัมพันธ์กับจำนวนประชากรลูกหลานที่จะเกิดขึ้นในการกำเนิดรุ่นต่อไป ดังสมการที่ (2.2)

$$F(x_i) = \frac{f(x_i)}{\sum_{i=1}^{Nind} f(x_i)} \quad (2.2)$$

โดย $F(x_i)$ คือ ค่าฟิตเนสของประชากรแต่ละตัวที่เกี่ยวข้องกับประชากรทั้งหมด $f(x_i)$ คือค่า ฟังก์ชันเป้าหมายของประชากรแต่ละตัว ซึ่ง N_{ind} คือจำนวนประชากรทั้งหมด และ x_i เป็นค่าฟีโนไทป์

การเลือกประชากรแต่ละตัวจากความสำเร็จจะเป็น แต่การเลือกอาจมีความผิดพลาดจากค่า ฟังก์ชันเป้าหมายของประชากรมีค่าเป็นลบ และอาจจะมีกรู้เข้าก่อนกำหนดจึงมีการแก้ปัญหาที่ เกิดขึ้น

2.4.1 สเกลลิงแบบเชิงเส้น(Linear Scaling)

วิธีนี้จะช่วยลดผลกระทบจากการรู้เข้าก่อนกำหนด แต่จะต้องระวังการเกิดค่าฟิตเนสที่ อาจจะเป็นลบจากการเลือกค่า a และ b จึงต้องพิจารณาขอบเขตของค่าฟิตเนสเดิมด้วย [2] ดัง สมการที่ (2.3)

$$F(x_i) = af(x_i) + b \quad (2.3)$$

โดยที่ $F(x_i)$ คือค่าฟิตเนสของแต่ละโครโมโซมที่การสเกลลิงแบบเชิงเส้น ที่มีความสัมพันธ์อย่าง เป็นเชิงเส้นกับ $f(x_i)$ ที่เป็นค่าฟังก์ชันเป้าหมาย และ a คือค่าของแฟคเตอร์สเกลที่มีค่าเป็นบวก b เป็นค่าชดเชย (Offset)

2.4.2 การตัดเศษซิกมา (Sigma Truncation)

วิธีนี้ใช้เพื่อหลีกเลี่ยงการเกิดฟิตเนสที่เป็นลบ ดังสมการที่ 2.4

$$F(x_i) = f(x_i) - (f(\bar{x}) - c\sigma) \quad (2.4)$$

เมื่อ \bar{x} คือ ค่าเฉลี่ยหรือมัธมิมเลขคณิต $F(x_i)$ คือค่าเฉลี่ยของฟังก์ชันเป้าหมาย σ คือส่วน เบี่ยงเบนมาตรฐานของประชากร และ c คือจำนวนเต็มที่มีค่าน้อย ๆ (ระหว่าง 1 ถึง 5)

2.4.3 สเกลลิงแบบยกกำลัง

วิธีนี้จะทำการยกกำลังฟังก์ชันเป้าหมาย ดังสมการที่ (2.5)

$$f(x_i) = f(x_i^k) \quad (2.5)$$

เมื่อ k เป็นค่าที่จะขึ้นอยู่กับปัญหาที่ต้องการแก้ไข หรือจะมีการเปลี่ยนแปลงค่าได้ในขณะที่มีการ รัน [6]

2.4.4 การกำหนดตำแหน่ง (Ranking)

วิธีนี้ช่วยหลีกเลี่ยงการลู่เข้าก่อนกำหนด และช่วยเพิ่มความเร็วในการค้นหา [7] โดยเป็นการเรียงค่าฟังก์ชันเป้าหมายที่มีค่าฟิตเนสน้อยสุดจะอยู่ในลำดับที่ 1 และจะเรียงลำดับต่อไปจะได้ค่าฟิตเนสที่มากที่สุดอยู่ลำดับสุดท้าย ดังสมการที่ (2.6)

$$Fitness(pos) = 2 - SP + 2(SP - 1) \frac{Pos - 1}{N_{ind} - 1} \quad (2.6)$$

โดย SP คือค่าความลำเอียง (Bias) หรือค่าของซีเล็กทีฟเพสเชอร์ (Selective Pressure) Pos คือลำดับของประชากรตัวที่ i และ N_{ind} คือจำนวนประชากรทั้งหมด

2.5 การคัดเลือก (Selection)

การคัดเลือกเป็นแบบจำลองการคัดเลือกทางธรรมชาติ ที่ใช้การเลือกสมาชิกแต่ละตัวในประชากรสำหรับกระบวนการรีโพรดักชัน โดยกลไกการคัดเลือกที่เป็นการจับคู่ประชากรพ่อแม่เป็นต้นแบบในบ่อเพาะพันธุ์ (Mating Pool) บนค่าฟิตเนสพื้นฐานสดริงของแต่ละประชากรสำหรับกระบวนการถ่ายทอดเพื่อสร้างประชากรรุ่นลูกหลานต่อไป การคัดเลือกในวิธีพันธุกรรมศาสตร์นั้น จะประกอบด้วยกำหนดตำแหน่ง การแข่งขัน และสัดส่วนตามทฤษฎี Schema [8] หรือทฤษฎีพื้นฐานของวิพันธุกรรมศาสตร์ โดยจะใช้ค่าความเหมาะสมเป็นตัวตัดสินว่าประชากรตัวไหนมีโอกาสดูกเลือกเป็นประชากรพ่อแม่ ซึ่งประชากรที่มีค่าเหมาะสมที่ดีจะถูกกำหนดค่าความน่าจะเป็นที่จะถูกเลือกแต่ละครั้ง จะสูงตามไปด้วยจากสมการที่ (2.7)

$$Pr ob_select = \frac{F_i}{\sum Fitness} \quad (2.7)$$

โดย F_i คือค่าฟิตเนสของประชากรแต่ละตัว และ $\sum Fitness$ คือผลรวมของค่าฟิตเนสทั้งหมดที่นำมาคัดเลือกในบ่อเพาะพันธุ์ ซึ่งมีมาตรการวัดการทำงานของวิธีการคัดเลือกมีดังนี้

Selective Pressure คือ ค่าความน่าจะเป็นของจำนวนประชากรที่ดีที่สุดที่ถูกเลือก โดยเปรียบเทียบกับค่าเฉลี่ยของความน่าจะเป็นของจำนวนประชากรที่ถูกเลือกทั้งหมด

Loss of Diversity คือ สัดส่วนของประชากรตัวหนึ่ง ที่ไม่ได้ถูกเลือกในระหว่างขั้นตอนของการคัดเลือก

ความหนาแน่นของการคัดเลือก (Selection Intensity) คือ ค่าฟิตเนสเฉลี่ยของประชากรหลังจากการใช้วิธีการคัดเลือกแบบการกระจายตัวแบบ Gaussian

ความแปรปรวนของการคัดเลือก (Selection Variance) คือ ความแปรปรวนของการกระจายตัวของค่าฟิตเนสของประชากร หลังจากการใช้วิธีการคัดเลือกแบบการกระจายตัวของ Gaussian

ความลำเอียง (Bias) คือ ค่าสัมบูรณ์ของค่าความแตกต่างระหว่างค่าจริงกับค่าความน่าจะเป็นที่คาดหวังไว้สำหรับการคัดเลือกสมาชิกประชากรแต่ละตัว

การกระจาย (Spread) คือ ค่าความน่าจะเป็นไปได้ที่ประชากรสมาชิกตัวหนึ่งจะถูกเลือกจากช่วงของจำนวนการทดลอง ถ้า $f(i)$ เป็นจำนวนของการทดสอบประชากรสมาชิกลำดับที่ i ที่มีการกระจายต่ำที่สุด เป็นการกระจายที่น้อยที่สุดย่อมมีความลำเอียงเป็นศูนย์ตามทฤษฎี

$$f(i) \in \{ \lfloor et(i) \rfloor, \lceil et(i) \rceil \} \quad (2.8)$$

เมื่อ $et(i)$ คือ จำนวนทดลองที่คาดหวังไว้ของประชากรสมาชิกลำดับที่ i , $\lfloor et(i) \rfloor$ คือ ขอบเขตล่าง และ $\lceil et(i) \rceil$ คือ ขอบเขตบน ขณะที่ความลำเอียงมีค่าแน่นอนแล้วการกระจายจะเป็นตัววัดความเที่ยงของวิธีการคัดเลือก โดยประสิทธิภาพ จะขึ้นอยู่กับความซับซ้อนของเวลาทั้งหมดสำหรับอัลกอริทึมนี้

2.5.1 การคัดเลือกแบบวงล้อรูเลต (Roulette Wheel)

วิธีการคัดเลือกแบบวงล้อรูเลตอย่างง่าย เป็นการเลือกแบบสุ่มโดยมีการแทนที่ (Stochastic Sampling with Replacment : SSR) [9] โดยขนาดของแต่ละช่องของวงล้อจะขึ้นอยู่กับค่าฟิตเนส เส้นรอบวงล้อจะมีค่าความน่าจะเป็นเท่ากับ 1 และจำนวนครั้งในการสุ่มขึ้นอยู่กับประชากรรุ่นพ่อแม่ที่ต้องการ วิธี SSR นี้ขนาดของเซกเมนต์และความน่าจะเป็นในการเลือกจะคงที่ตลอดช่วงการคัดเลือก ซึ่งมีแนวโน้มให้ความลำเอียงเป็นศูนย์แต่การกระจายไม่ได้ให้ค่าน้อยที่สุด

การคัดเลือกแบบสุ่มโดยมีการแทนที่บางส่วน (Stochastic Sampling with Partial Replacement : SSPR) เป็นวิธีที่มีการปรับเปลี่ยนขนาดของเซกเมนต์ของโครโมโซม โดยโครโมโซมที่มีการถูกเลือกในแต่ละครั้งจะมีการลดขนาดของเซกเมนต์นั้นด้วยแฟกเตอร์ค่าหนึ่ง ซึ่งถ้าขนาดเซกเมนต์ของโครโมโซมมีค่าเป็นลบ ขนาดของเซกเมนต์นั้นจะถูกกำหนดให้เป็นศูนย์ วิธีนี้จะกำหนดขอบเขตบนของการกระจาย $\lceil et(i) \rceil$ และขอบเขตล่างมีค่าเป็นศูนย์ ซึ่งจะทำให้มีความลำเอียงที่สูงขึ้น วิธีการคัดเลือกแบบวงล้อรูเลตมีความซับซ้อนของเวลาเป็น $N \log N$ เมื่อ N คือ ขนาดของประชากร

2.5.2 การคัดเลือกแบบยูนิเวอร์ซอลโดยสุ่ม (Stochastic Universal Sampling : SUS)

วิธีการคัดเลือกที่มีความลำเอียงเท่ากับศูนย์ และจะได้ค่าการกระจายน้อยที่สุด ซึ่งวิธีนี้เป็น การแก้ปัญหาของวิธีการคัดเลือกแบบวงล้อรูเลต ดังนั้นในแต่ละช่องของโครโมโซมจะเหมือนกับวิธี

คัดเลือกแบบวงล้อรูเล็ต โดยจะเพิ่มตัวชี้ N-pointer ที่มีจำนวนเท่ากับจำนวนโครโมโซมที่ต้องการเลือกซึ่งมีระยะห่างของตัวชี้เท่ากันเป็น 1/N-pointer โดยตำแหน่งแรกของตัวชี้จะกำหนดจากการสุ่มในช่วง $[0, 1/N\text{-pointer}]$

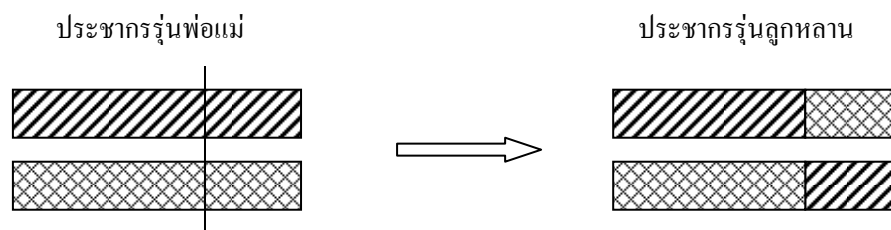
2.6 ตัวดำเนินการพันธุกรรมศาสตร์ (Genetic Operator)

2.6.1 การรวมตัวของยีน (Recombination)

เป็นการเพิ่มประสิทธิภาพให้แก่วิธีพันธุกรรมศาสตร์ ที่จะค้นหาสิ่งที่ดีกว่าในโครโมโซมฝ่ายพ่อและฝ่ายแม่ ที่ได้จากการคัดเลือกแล้วที่มีอยู่ในบ่อเพาะพันธุ์ มาทำการจับคู่ถ่ายทอดยีนของประชากรพ่อแม่ ตามอัตราความน่าจะเป็นของการรวมตัวของยีนหรือการครอสโอเวอร์ (Probability of Crossover : Pc) เพื่อสร้างประชากรรุ่นใหม่ โดยจะแบ่งการรวมตัวของยีนนี้เป็น 2 ประเภท คือ การรวมตัวของยีนโดยใช้ค่าไบนารี และการรวมตัวของยีนโดยใช้ค่าจริง

2.6.1.1 การรวมตัวของยีนโดยใช้ค่าไบนารี (Binary Valued Recombination)

- การครอสโอเวอร์แบบจุดเดียว (Single point Crossover) จะทำการครอสโอเวอร์หนึ่งจุดในช่วง $[1, 2, \dots, Nvar-1]$ โดยที่ Nvar คือจำนวนตัวแปรของแต่ละโครโมโซม หรือ ประชากร ดังรูปที่ 2.3 แสดงการครอสโอเวอร์แบบจุดเดียว



รูปที่ 2.3 การครอสโอเวอร์แบบจุดเดียว

พิจารณาสตริงไบนารีของประชากรพ่อแม่ 2 ตัว โดยประชากรพ่อแม่แต่ละตัวจะมีตัวแปร 10 ตัว โดยตำแหน่งของการครอสโอเวอร์ที่สุ่มได้เท่ากับ 7

ประชากรพ่อแม่ 1	1	0	1	0	0	1	1	0	1	1
ประชากรพ่อแม่ 2	0	1	0	1	0	0	1	1	0	1
ประชากรลูกหลาน 1	1	0	1	0	0	1	1	1	0	1
ประชากรลูกหลาน 2	0	1	0	1	0	0	1	0	1	1

ประชากรลูกหลานที่ได้ทั้งคู่นั้นจะมีความแตกต่างกับประชากรพ่อแม่ ซึ่งถ้าประชากรพ่อแม่มีค่าฟิตเนสที่สูงจะถูกสงวนไว้ในการกำเนิดถัดไป แล้วเป็นส่วนย่อยที่ประชากรลูกหลานที่ได้

นั้นจะมีความเหมาะสมกว่าประชากรพ่อแม่ เพราะการครอสโอเวอร์มีแนวโน้มอนุรักษ์ข้อมูลทางพันธุกรรมศาสตร์ที่มีในโครโมโซมฝ่ายพ่อและฝ่ายแม่ จากการครอสโอเวอร์แบบจุดมีข้อเสียในบางกรณีที่สคีมาบางตัวไม่สามารถรวมกันได้ เช่น

$$\begin{aligned} S1 &= 1 \quad 0 \quad 1 \quad * \quad * \quad * \quad * \quad 1 \\ S2 &= * \quad * \quad * \quad * \quad 1 \quad 1 \quad * \quad * \end{aligned}$$

โดยที่ S1 , S2 เป็นสคีมา และ “ * ” มีสัญลักษณ์ “don't care” มีโครโมโซม 2 ตัว คือ C1 และ C2 ที่ตรงกับ S1 และ S2 ตามลำดับ

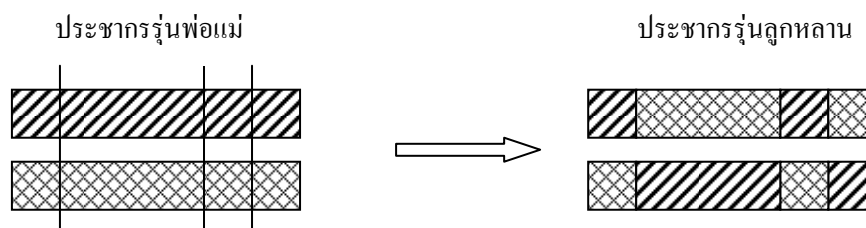
$$\begin{aligned} C1 &= 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \\ C2 &= 0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \end{aligned}$$

ถ้าทำการครอสโอเวอร์แบบจุดเดียว จะไม่มีทางได้โครโมโซมที่ตรงกับสคีมา S3

$$S3 = 1 \quad 0 \quad 1 \quad * \quad 1 \quad 1 \quad * \quad 1$$

ซึ่งสามารถแก้ปัญหานี้ได้โดยใช้การครอสโอเวอร์แบบหลายจุดที่มีการเพิ่มความสามารถในการสร้างโครโมโซมลูกหลาน

- การครอสโอเวอร์แบบหลายจุด (Multi – point Crossover) การครอสโอเวอร์อยู่ในช่วง [1,2,..., Nvar-1] โดยที่ Nvar คือจำนวนตัวแปรของประชากรแต่ละตัว ซึ่งจะมีการเลือกจุดที่จะทำการครอสโอเวอร์โดยการสุ่มตามจำนวน m ต่อจากนั้นตัวแปรจะทำการแลกเปลี่ยนกันระหว่างประชากรพ่อแม่ทั้งสอง โดยจะไม่มีแลกเปลี่ยนกันระหว่างตัวแปรในตำแหน่งแรกของการครอสโอเวอร์ แสดงในรูปที่ 2.4



รูปที่ 2.4 การครอสโอเวอร์แบบหลายจุด

จาก C1 และ C2 ถ้าการครอสโอเวอร์แบบ 2 จุด ตำแหน่งที่สุ่มคือตำแหน่งที่ 3 และ 5 ซึ่งจะทำให้ได้ C3 และ C4 ที่เป็นโครโมโซมลูกหลานที่ C3 จะตรงกับสคีมา S3

$$\begin{aligned}
 C1 &= 1 & 0 & 1 & 1 & | & 0 & 0 & | & 0 & 1 \\
 C2 &= 0 & 1 & 1 & 0 & | & 1 & 1 & | & 0 & 0 \\
 C3 &= 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & \text{ตรงกับสตีมา S3} \\
 C4 &= 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0
 \end{aligned}$$

- การครอสโอเวอร์แบบยูนิฟอร์ม (Uniform Crossover) การครอสโอเวอร์แบบยูนิฟอร์ม [10] ซึ่งเป็นการสร้างประชากรรุ่นลูกหลาน โดยการสุ่มของการสร้างระบุตำแหน่ง (Mask) พิจารณาจาก

ประชากรพ่อแม่ 1	1	0	1	0	0	1	1	1	0	0	1
ประชากรพ่อแม่ 2	0	1	0	1	1	0	1	0	1	1	0
Mask	0	1	0	1	1	0	0	1	1	0	1
ประชากรลูกหลาน 1	0	0	0	0	0	0	1	1	0	1	1
ประชากรลูกหลาน 2	1	1	1	1	1	1	1	0	1	0	0

การครอสโอเวอร์แบบยูนิฟอร์มเป็นการแลกเปลี่ยนข้อมูลระหว่างบิต ของประชากรพ่อแม่ ก็ต่อเมื่อบิต ของ Mask ที่ตำแหน่งเป็น 0 และจะไม่สลับข้อมูลเมื่อบิต Mask เป็นตำแหน่งที่ 1

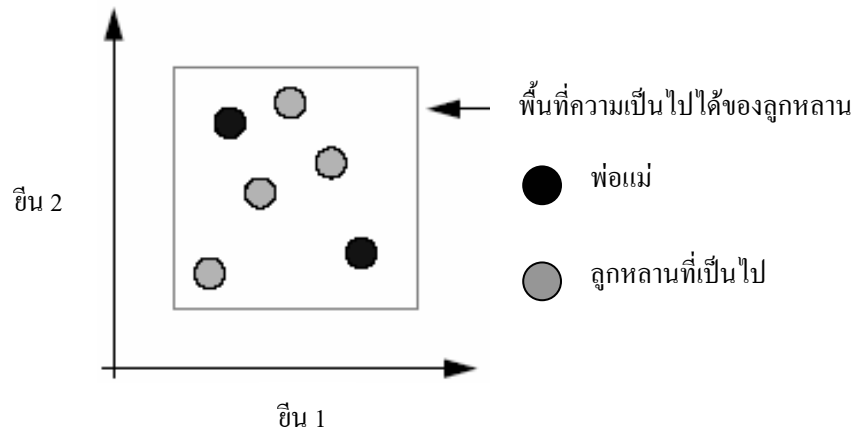
2.6.1.2 การรวมตัวของยีนโดยใช้ค่าจริง

- การรวมตัวของยีนแบบไม่ต่อเนื่อง (Discrete Recombination) เป็นการปรับเปลี่ยนข้อมูล ค่าตัวแปรที่เป็นค่าจำนวนจริงของประชากรรุ่นพ่อแม่ โดยจะทำการใช้ข้อมูลตัวแปรไหนขึ้นอยู่กับ ค่าความน่าจะเป็นที่สุ่มได้ และจะปรับเปลี่ยนยีนที่อยู่ตำแหน่งเดียวกันเท่านั้น จะทำให้ได้ประชากร รุ่นลูกหลาน

- การรวมตัวของยีนแบบขั้นกลาง [11] (Intermediate Recombination) ใช้กับตัวแปรที่ เข้ารหัสแบบค่าจริงเท่านั้น โดยประชากรรุ่นลูกหลานที่เกิดขึ้นจากสมการที่ (2.9)

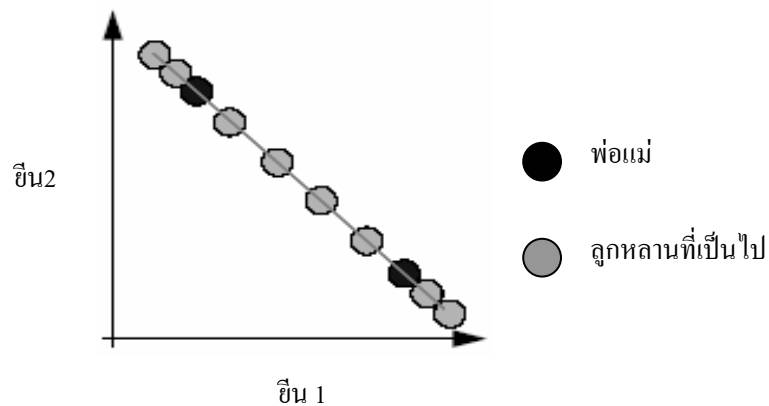
$$\text{ประชากรรุ่นลูกหลาน} = \text{ประชากรพ่อแม่ 1} + \alpha (\text{ประชากรพ่อแม่ 2} - \text{ประชากรพ่อแม่ 1}) \quad (2.9)$$

โดยที่ α เป็นแฟกเตอร์สเกล (scaling factor) ที่ถูกสุ่มเลือกโดยมีค่าอยู่ระหว่าง $[-d, 1+d]$ ซึ่ง ขอบเขตของประชากรพ่อแม่ที่อยู่นั้นอยู่ระหว่าง 0 ถึง 1 ส่วน d จะเป็นตัวขยายขอบเขตของประชากร รุ่นลูกหลานที่มีค่ามากกว่าค่าของข้อมูลตัวแปรประชากรพ่อแม่เท่ากับค่าของ α ที่ถูกเลือก ค่าที่ดี ที่สุดที่ถูกเลือกจะมีค่า $d=0.25$ การรวมตัวของยีนแบบขั้นกลางสามารถสร้างจุดภายในไฮเปอร์คิวบ์ (Hypercube) โดยที่ความกว้างจะขึ้นอยู่กับตัวประชากรพ่อแม่ซึ่งเป็นตัวกำหนด รูปที่ 2.5 แสดง ขอบเขตการกำเนิดของประชากรรุ่นลูกหลาน



รูปที่ 2.5 ขอบเขตการกำเนิดของประชากรรุ่นลูกหลานแบบการรวมตัวยีนแบบชั้นกลาง

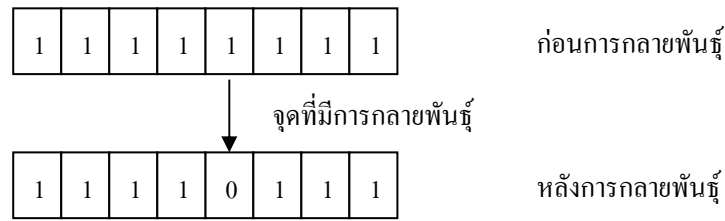
- การรวมตัวของยีนแบบเส้น (Line Recombination) เป็นวิธีที่จะใช้ค่า α ที่สุ่มได้ค่าเดียวกับค่าตัวแปรของประชากรทั้งหมด โดยประชากรรุ่นลูกหลานที่เกิดขึ้นจะเรียงเป็นเส้นที่ถูกกำหนดโดยประชากรพ่อแม่ ดังแสดงในรูปที่ 2.6



รูปที่ 2.6 การรวมตัวของยีนแบบเส้น

2.6.2 การกลายพันธุ์ (Mutation)

การกลายพันธุ์เป็นการสร้างความหลากหลายให้กับประชากร โดยจะทำการสุ่มเปลี่ยนข้อมูลทางพันธุกรรมศาสตร์หรือข้อมูลตัวแปรในประชากร ซึ่งอาจจะทำให้ได้ประชากรที่ดีขึ้นหรือแย่ลง ด้วยเหตุที่การกลายพันธุ์นั้นจะมีความแตกต่างกันในประชากร และมีการค้นหาที่นอกเหนือจากการครอสโอเวอร์ จึงอาจจะทำให้เจอตําแหน่งของประชากรที่ดีกว่าประชากรเก่า ซึ่งแสดงลักษณะการกลายพันธุ์ในรูปที่ 2.7



รูปที่ 2.7 การกลายพันธุ์

โดยทั่วไปแล้วการกลายพันธุ์จะออกแบบไว้สำหรับประชากรที่เข้ารหัสแบบไบนารีเท่านั้น แต่การกลายพันธุ์แบบสุ่ม [12] เป็นวิธีหนึ่งที่มีการใช้กันในกรณีที่ประชากรเป็นแบบจำนวนจริง เป็นไปตามสมการที่ (2.10)

$$g = g + \psi(\mu, \sigma) \quad (2.10)$$

เมื่อ g คือ ยีนที่เป็นค่าจริง ψ คือ ฟังก์ชันสุ่ม ซึ่งอาจจะเป็นแบบ Gaussian หรือ การสุ่มแบบกระจายปกติ (Normal Distribution) μ, σ คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนของฟังก์ชันสุ่มตามลำดับ

2.7 การแทนที่ (Reinsertion)

เมื่อผ่านตัวดำเนินการพันธุกรรมศาสตร์ จะได้ประชากรรุ่นลูกหลานซึ่งจะมีการแทนที่ในประชากรพ่อแม่ปัจจุบันด้วย ประชากรรุ่นลูกหลานที่เกิดขึ้นทุกตัว จึงทำให้การแทนที่นี้อาจทำให้เกิดความเสียหายกับประชากรในการกำเนิดครั้งต่อไป ดังนั้นจึงต้องมีการทำสำเนาประชากรที่ดีที่สุด 2-5 ตัว ในการกำเนิดครั้งถัดไป ในการทำสำเนาประชากรนี้จะช่วยเพิ่มความเร็วในการหาคำตอบที่ดีที่สุด ถ้าหากกำหนดจำนวนสำเนาประชากรได้อย่างเหมาะสม

เนื่องจากในการสร้างประชากรรุ่นลูกหลานที่มีขนาดใหญ่ จะทำให้ใช้เวลานานขึ้นในการคำนวณในแต่ละรอบของวิพันธุกรรมศาสตร์ จึงมีวิธีการแทนที่แบบอื่นที่สร้างประชากรลูกหลานให้มีจำนวนน้อยลง เพื่อช่วยเพิ่มความเร็วในการคำนวณ เช่น การแทนที่ประชากรพ่อแม่ด้วยประชากรลูกหลานโดยตรง คือจะเลือกแทนที่ประชากรที่เลวที่สุดเมื่อมีการแทรกประชากรรุ่นใหม่เข้ามา หรืออาจแทนที่ประชากรที่อาศัยอยู่เป็นเวลานาน

คุณลักษณะวิพันธุกรรมศาสตร์ในบทที่ 2 จะถูกนำไปใช้ประยุกต์กับการจัดสรรเงินลงทุน แต่การจัดสรรเงินลงทุนในกลุ่มหลักทรัพย์ที่ดีต้องจัดความเสี่ยง และการคำนวณค่าความเสี่ยงที่เกิดขึ้นได้ โดยในบทที่ 3 จะกล่าวถึงเรื่องค่าความเสี่ยง