

ผลกระทบของขนาดต้นไม้เริ่มต้นที่มีต่อค่าความเพียรพยายามเชิงคำนวณของกำหนดการเชิงพันธุกรรม

Impact of the Initial Tree Size on Computational Effort in Genetic Programming

ขนิษฐา หุตะเมขลิน นัทธี นิปานันท์ ประภาส จงสถิตย์วัฒนา

Kanitha Huthameklin, Nattee Niparnan, Prabhas Chongstitvatana

Department of Computer Engineering

Chulalongkorn University

Bangkok 10300, Thailand

บทคัดย่อ

Genetic Programming has been widely used to automatically generate program to solve various problems. However, GP usually requires substantial computational effort. In this work, we study the impact of the initial tree size on computational effort in GP by using robotic arms moving to targets as the case study. The experimental design and hypothesis testing has been done by statistical approach. After the data has been analyzed using the analysis of variance technique, the result shows that the initial tree size effects the computational effort. By using the pair comparison under the level of significance 0.05, the result also indicates that computational effort is decreased when the initial tree size is larger.

Key words: Computational Effort, Genetic Programming.

กำหนดการเชิงพันธุกรรมเป็นกระบวนการที่ถูกนำไปใช้ในการสร้างโปรแกรมเพื่อแก้ปัญหาต่าง ๆ อย่างแพร่หลาย อย่างไรก็ตามตามกำหนดการเชิงพันธุกรรมต้องการความเพียรพยายามเชิงคำนวณเป็นอย่างมากในการทำงาน บทความนี้จะศึกษาถึงผลกระทบของขนาดต้นไม้เริ่มต้นที่มีต่อค่าความเพียรพยายามเชิงคำนวณของกำหนดการเชิงพันธุกรรม โดยใช้ปัญหาการเคลื่อนที่ของแขนหุ่นยนต์ไปยังเป้าหมายเป็นกรณีศึกษา การศึกษาผลกระทบดังกล่าวใช้กระบวนการทางสถิติในการวางแผนการทดลองและทดสอบสมมติฐาน จากการทดลองเมื่อวิเคราะห์ข้อมูลด้วยเทคนิควิเคราะห์ความแปรปรวนโดยกำหนดระดับนัยสำคัญของการทดสอบสมมติฐานเท่ากับ 0.05 พบว่าขนาดของต้นไม้เริ่มต้นมีผลกระทบต่อค่าความเพียรพยายามเชิงคำนวณ และเมื่อใช้วิธีการทดสอบสมมติฐานเกี่ยวกับผลต่างระหว่างค่าเฉลี่ยของสองประชากรแบบจับคู่ กำหนดระดับนัยสำคัญรวมเท่ากับ 0.05 จะพบว่าเมื่อขนาดของต้นไม้เริ่มต้นใหญ่ขึ้น จะทำให้ค่าความเพียรพยายามเชิงคำนวณเฉลี่ยลดลง

1. บทนำ

กำหนดการเชิงพันธุกรรมเป็นกระบวนการหนึ่งในการสร้างโปรแกรมสำหรับควบคุมการทำงานต่าง ๆ ขึ้นมาโดยอัตโนมัติ หน่วยวัดประสิทธิภาพของกำหนดการเชิงพันธุกรรมวิธีหนึ่งที่ใช้คือค่าความเพียรพยายามในการคำนวณ Koza [5] ได้กำหนดนิยามของค่าความเพียรพยายามเชิงคำนวณไว้ว่า เป็นจำนวนของผลเฉลยน้อยที่สุดที่ต้องประมวลเพื่อให้ได้คำตอบที่ต้องการ

ชัยวัฒน์ เจษฎาปกรณ์ [1] ได้วิจัยการลดทอนค่าความเพียรพยายามเชิงคำนวณของวิธีการเรียนรู้แบบกำหนดการเชิงพันธุกรรม โดยใช้ปัญหาแขนหุ่นยนต์เคลื่อนที่ไปยังวัตถุเป้าหมาย และได้ให้ข้อสังเกตไว้ว่า “จากผลการทดลองการหาความสัมพันธ์ระหว่างขนาดของโปรแกรมหุ่นยนต์เริ่มต้นและค่าความเพียรพยายามเชิงคำนวณ พบว่าขนาดของผลเฉลยโปรแกรมหุ่นยนต์เริ่มต้นมีผลต่อค่าความเพียรพยายามเชิงคำนวณเป็นอย่างมาก ซึ่งไม่สามารถอธิบายได้ว่าเป็นเพราะเหตุใด”

ในบทความนี้ จะศึกษาถึงผลกระทบของขนาดของโปรแกรมเริ่มต้นที่มีต่อค่าความเพียรพยายามเชิงคำนวณว่ามีหรือไม่และเป็นไปในลักษณะใด และเมื่อลักษณะของปัญหาเปลี่ยนไป หรือเมื่อค่าพารามิเตอร์ต่าง ๆ ของกำหนดการเชิงพันธุกรรมเปลี่ยนไป ผลกระทบดังกล่าวยังมีอยู่หรือไม่ โดยใช้ปัญหาแขนหุ่นยนต์เคลื่อนที่ไปยังเป้าหมายเป็นกรณีศึกษา การตอบคำถามดังกล่าวข้างต้น จะวิเคราะห์โดยใช้หลักการทางสถิติ

2. งานที่เกี่ยวข้อง

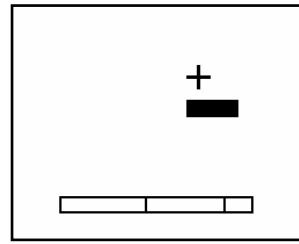
กรณีศึกษาในบทความนี้ ได้นำมาจากงานวิจัยเรื่อง การลดทอนความเพียรพยายามเชิงคำนวณของวิธีการเรียนรู้แบบกำหนดการเชิงพันธุกรรม ซึ่งในงานวิจัยดังกล่าวได้นิยามปัญหาแขนหุ่นยนต์เคลื่อนที่ไปยังเป้าหมาย และกำหนดรายละเอียดของกำหนดการเชิงพันธุกรรม

ในบทความ [2] ดันไม้ 1 ดันจะประกอบด้วยส่วนเด่นและส่วนซ่อนเร้น ถ้าจำนวนโหนดที่เป็นส่วนซ่อนเร้นมีมากกว่าจะทำให้โอกาสที่การไขว้เปลี่ยนจะทำลายส่วนเด่นนั้นลดลง และสามารถถ่ายทอดส่วนเด่นที่ดีไปยังดันไม้ในรุ่นต่อไปได้ บทความ [3] กล่าวว่า เมื่อขนาดของดันไม้มีขนาดใหญ่ขึ้น จะมีส่วนของดันไม้ที่เป็นส่วนซ่อนเร้นมากขึ้นด้วย ซึ่งทำให้โอกาสที่การไขว้เปลี่ยนจะทำลายส่วนเด่นของดันไม้ลดลง และส่วนซ่อนเร้นดังกล่าว อาจจะถูกกลายมาเป็นส่วนเด่นได้ในภายหลัง เนื่องจากการไขว้เปลี่ยนในรุ่นต่อ ๆ มา บทความ [4] กล่าวว่า การเพิ่มขึ้นของขนาดดันไม้มีความเกี่ยวข้องกับโอกาสที่ดันไม้ในรุ่นต่อไปจะมีค่าความเหมาะสมดีขึ้น ดันไม้ที่มีขนาดใหญ่จะถูกเลือกเพื่อทำให้ค่าความเหมาะสมดีขึ้น

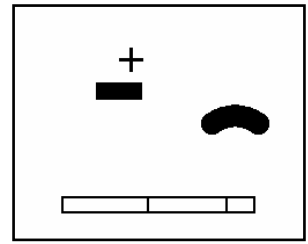
บทความทั้งสาม ใช้ปัญหาที่มีลักษณะของโปรแกรมขึ้นอยู่ กับตำแหน่งของฟังก์ชันและเทอมินอล คือ โปรแกรมที่ประกอบด้วยทดสอบเงื่อนไข if-then-else ซึ่งเป็นโปรแกรมลักษณะเดียวกับที่ใช้ในกรณีศึกษาของบทความนี้ บทความนี้สนใจเจาะจงที่ขนาดของดันไม้

3. ปัญหาการควบคุมแขนหุ่นยนต์

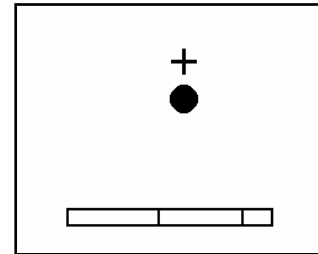
ปัญหาดังกล่าวเป็นการควบคุมให้แขนของหุ่นยนต์เคลื่อนที่ไปยังเป้าหมายที่กำหนดไว้ โดยหลีกเลี่ยงสิ่งกีดขวาง บทความนี้ได้ใช้ระบบจำลองในการทดลอง ระบบจำลองประกอบด้วย แขนหุ่นยนต์ 3 ท่อน ได้แก่ท่อนไหล่ ท่อนแขน และท่อนมือ เชื่อมต่อกันด้วยข้อต่อ และมีข้อต่อที่เชื่อมระหว่างท่อนไหล่กับพื้น ข้อต่อแต่ละข้อต่อสามารถหมุนในแนวแนวระนาบอย่างเป็นอิสระต่อกัน โดยมีมุมการเคลื่อนไหวที่จำกัด สภาพแวดล้อมที่ใช้ในการทดลองมีทั้งหมด 3 สภาพแวดล้อม โดยแต่ละสภาพแวดล้อมมีขนาด 200 x 200 จุดภาพ ดังแสดงดังรูปต่อไปนี้



รูปที่ 1 : สภาพแวดล้อมที่ 1



รูปที่ 2 : สภาพแวดล้อมที่ 2



รูปที่ 3 : สภาพแวดล้อมที่ 3

เป้าหมายของสภาพแวดล้อมแสดงด้วยเครื่องหมาย + สิ่งกีดขวางแสดงด้วยพื้นที่ที่ทึบสีดำ และแขนหุ่นยนต์แทนด้วยสี่เหลี่ยม 3 ท่อนที่ต่อกันอยู่ด้านล่าง

4. การทดลอง

เราต้องการทดลองเพื่อหาผลกระทบของขนาดดันไม้เริ่มต้น (ปัจจุัย) ที่มีผลต่อค่าความเพียรพยายามเชิงคำนวณ (ตัวแปรตาม) โดยพยายามตัดผลกระทบที่มีต่อค่าความเพียรพยายามเชิงคำนวณที่เกิดจากค่าพารามิเตอร์อื่นๆ (ปัจจุัยรบกวน) ในบทความนี้ เราใช้แผนการทดลองแบบสุ่มสมบูรณ์ภายในกลุ่ม (Randomized complete block design) เป็นวิธีการกำจัดผลกระทบของปัจจุัยรบกวนที่มีต่อตัวแปรตาม

แผนการทดลองแบบสุ่มสมบูรณ์ภายในกลุ่มเป็นการแบ่งกลุ่มของหน่วยการทดลองตามปัจจุัยรบกวนต่าง ๆ และสุ่มเลือกหน่วยทดลองภายในแต่ละกลุ่มให้กับระดับปัจจุัยที่ต้องการศึกษา โดยแต่ละกลุ่มต้องแตกต่างกันอย่างชัดเจน เพื่อให้หน่วยทดลองในกลุ่มเดียวกันมีความคล้ายคลึงกันมากที่สุด การแบ่งกลุ่มนั้นจะช่วยขจัดความผันแปรที่เกิดจากปัจจุัยรบกวน และช่วยให้ควบคุมปัจจุัยรบกวนได้อย่างเป็นระบบ

เกณฑ์ในการกำหนดระดับและขอบเขตของดันไม้เริ่มต้นได้มาจากข้อมูลในงานวิจัย [1] โดยกำหนดขอบเขตของขนาดดันไม้เริ่มต้นตั้งแต่ 40 ถึง 400 โหนด และความแตกต่างระหว่างระดับของขนาดดันไม้เริ่มต้นเท่ากับ 10 โหนด

พารามิเตอร์ต่าง ๆ จะถูกเลือกมาจัดเป็นชุดของพารามิเตอร์ ในการทดลองนี้ ได้เลือกพารามิเตอร์ทั้งหมดจำนวน 105 ชุด โดยจะใช้พารามิเตอร์แต่ละชุดในการประมวลผล กำหนดการเชิงพันธุกรรม ภายใต้สภาพแวดล้อมทั้ง 3

สภาพแวดล้อม ขอบเขตและระดับของพารามิเตอร์ที่ถูกเลือก จะแสดงดังตารางต่อไปนี้

ตาราง 2: ค่าพารามิเตอร์ที่ใช้ในการแบ่งกลุ่ม

พารามิเตอร์	ขอบเขต	ค่า
จำนวนประชากร	100 – 400	100,150,200,250,300,350,400
อัตราการสืบพันธุ์	5% - 10%	5%, 8%, 10%
อัตราการไขว้เปลี่ยน	40% - 60%	40%,45%,50%,55%,60%
อัตราการกลายแบบต่อยอค และ อัตราการกลายแบบต่อปลาย	รวมกันเท่ากับ 30% - 55%	(100% - อัตราการสืบพันธุ์ – อัตราการไขว้เปลี่ยน) / 2
สภาพแวดล้อมของแขนหุ่นยนต์	สภาพแวดล้อม 1 – 3	สภาพแวดล้อม 1 – 3

5. การวิเคราะห์ผลการทดลอง

ในบทความนี้ต้องการศึกษาถึงผลกระทบของขนาดของโปรแกรมเริ่มต้นที่มีต่อค่าความเพียรพยายามเชิงคำนวณใน 2 ประเด็น ประเด็นแรกคือขนาดของโปรแกรมเริ่มต้นมีผลต่อค่าความเพียรพยายามหรือไม่ ประเด็นที่ 2 คือขนาดของโปรแกรมเริ่มต้นมีผลกระทบต่อค่าความเพียรพยายามในลักษณะใด

5.1 การทดสอบว่าขนาดของโปรแกรมเริ่มต้นมีผลต่อค่าความเพียรพยายามเชิงคำนวณหรือไม่

การทดสอบสมมติฐานจะใช้วิธีการวิเคราะห์ความแปรปรวน (Analysis of variance) ซึ่งเป็นวิธีการที่นิยมใช้ในการทดสอบความแตกต่างของค่าเฉลี่ยของประชากร โดยทำการแบ่งค่าความแปรปรวนของข้อมูลออกเป็น ส่วน ๆ ตามแหล่งที่มาของความแปรปรวน แล้วตรวจสอบว่าค่าความแปรปรวนที่เกิดจากระดับของปัจจัยมีค่าใกล้เคียงกับความแปรปรวนที่เกิดจากความคลาดเคลื่อนในการทดลองหรือไม่

ในการทดสอบนี้จะมีการกำหนดโมเดลทางสถิติเพื่ออธิบายว่าค่าสังเกตมีส่วนประกอบอะไรบ้าง โดยโมเดลที่ใช้เป็นแบบ fixed effects model ซึ่งเป็นโมเดลที่สนใจความแตกต่างของประชากรที่ได้รับระดับของปัจจัยเฉพาะระดับของปัจจัยที่เลือกมาศึกษาผลกระทบเท่านั้น โมเดลแบบ fixed effects model มีลักษณะคือ $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ โดยที่ μ คือค่าเฉลี่ยโดยทั่วไป (general grand mean)

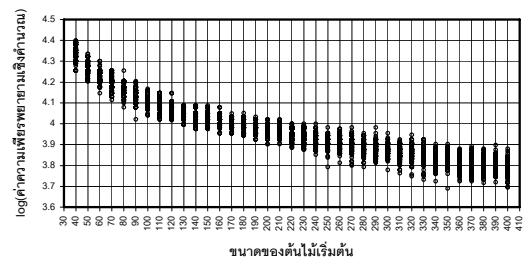
τ_i คือค่าแสดงผลที่ได้รับจากปัจจัยระดับที่ i

β_j คือค่าแสดงผลที่ได้รับจากกลุ่มที่ j

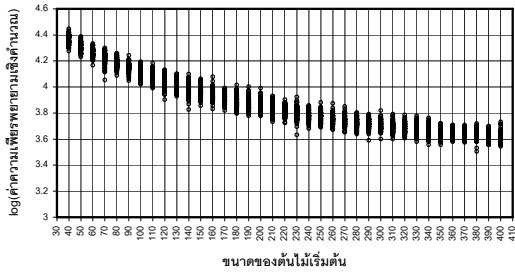
ε_{ij} คือค่าความคลาดเคลื่อน

ในโมเดลแบบ fixed effects model จะมีข้อสมมติคือแต่ละ ε_{ij} เป็นอิสระต่อกัน และมีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 0 และค่าความแปรปรวน σ^2 แต่จากผลการทดลองจะพบว่า ค่าที่ค่าความเพียรพยายามเชิงคำนวณมีการกระจายไม่เท่ากันในแต่ละขนาดของต้นไม้อื่น ดังนั้นจึงทำการแปลงข้อมูลเพื่อให้ข้อมูลมีลักษณะใกล้เคียงกับข้อสมมติมากที่สุด การแปลงข้อมูลที่ใช้คือการใช้ log โดยทำการหาค่า log ฐาน 10 ของข้อมูลแต่ละตัว ก่อนนำข้อมูลมาใช้

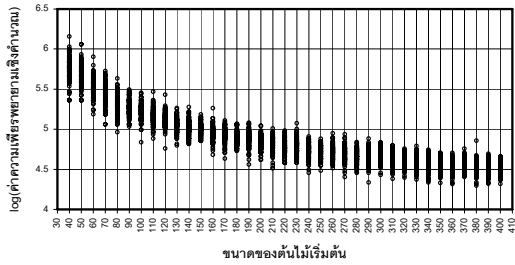
ผลการวิเคราะห์ข้อมูลทั้ง 3 สภาพแวดล้อมสรุปได้ว่าขนาดของต้นไม้อื่นที่ต่างกันมีผลทำให้ค่าความเพียรพยายามเชิงคำนวณเฉลี่ยต่างกันที่ระดับความเชื่อมั่น 95%



รูปที่ 4 : ค่าความเพียรพยายามเชิงคำนวณในสภาพแวดล้อมที่ 1



รูปที่ 5 : ค่าความฟิตรพยายามเชิงคำนวณในสภาพแวดล้อมที่ 2



รูปที่ 6 : ค่าความฟิตรพยายามเชิงคำนวณในสภาพแวดล้อมที่ 3

5.2 การทดสอบว่าขนาดของโปรแกรมเริ่มต้นที่เพิ่มขึ้นทำให้ค่าความฟิตรพยายามเชิงคำนวณลดลงหรือไม่

เราตั้งสมมติฐานในการวิจัยไว้ว่า เมื่อกำหนดให้ขนาดต้นไม้เริ่มต้นมีขนาดเพิ่มขึ้นจะทำให้ค่าความฟิตรพยายามเชิงคำนวณลดลง ซึ่งสมมติฐานนี้เขียนเป็นสมมติฐานทางสถิติได้คือ

$$H_0 : \mu_d = 0 \quad ; \quad \mu_d = \mu_i - \mu_j \quad \text{และ} \quad i < j$$

$$H_1 : \mu_d > 0 \quad ; \quad \mu_d = \mu_i - \mu_j \quad \text{และ} \quad i < j$$

โดยจะต้องทำการทดสอบสมมติฐานทั้งหมด

$\binom{37}{2} = 666$ คู่ในแต่ละสภาพแวดล้อม กำหนดระดับนัยสำคัญเท่ากับ α ทารด้วยจำนวนคู่ของสมมติฐาน

เนื่องจากเราต้องการกำจัดอิทธิพลของตัวแปรอื่นที่มีผลต่อค่าความฟิตรพยายามเชิงคำนวณ ดังนั้น เราจึงจับคู่ของประชากรในกลุ่มพารามิเตอร์เดียวกัน แล้วหาค่าความแตกต่าง d_k ของแต่ละคู่ระดับของปัจจัย แล้วจึงนำค่าความแตกต่างที่ได้มาคำนวณค่าสถิติทดสอบ

ตารางที่ 4: ผลการทดสอบสมมติฐาน

สภาพแวดล้อม	ผลการทดสอบสมมติฐาน	
	ปฏิเสธ H_0	ยอมรับ H_0
สภาพแวดล้อมที่ 1	637	29
สภาพแวดล้อมที่ 2	640	26
สภาพแวดล้อมที่ 3	618	48

จากผลการทดลอง สรุปได้ว่าประมาณ 90% ของจำนวนสมมติฐานทั้งหมดที่ได้ทำการทดสอบในแต่ละสภาพแวดล้อมขนาดของต้นไม้เริ่มต้นที่ใหญ่กว่ามีค่าความฟิตรพยายามเชิงคำนวณเฉลี่ยน้อยกว่าขนาดของต้นไม้เริ่มต้นที่เล็กกว่า

6. สรุป

บทความนี้เป็นการศึกษาผลกระทบของขนาดของโปรแกรมเริ่มต้นที่มีต่อค่าความฟิตรพยายามเชิงคำนวณ โดยตั้งสมมติฐานว่า ขนาดของโปรแกรมเริ่มต้นที่ใหญ่ขึ้น จะทำให้ค่าความฟิตรพยายามเชิงคำนวณลดต่ำลง และได้ใช้กระบวนการทางสถิติมาใช้ในการทดสอบสมมติฐานดังกล่าว การทดสอบสมมติฐานดังกล่าวกระทำภายใต้ความเชื่อมั่นทางสถิติ 95 % พบว่าประมาณ 90 % ของสมมติฐานที่ทำการทดสอบทั้งหมดขนาดของต้นไม้เริ่มต้นที่ใหญ่ขึ้นจะทำให้ค่าความฟิตรพยายามเชิงคำนวณเฉลี่ยลดลง

อย่างไรก็ตาม การกำหนดขนาดของต้นไม้เริ่มต้นให้ใหญ่นั้นอาจไม่เหมาะสมในทางปฏิบัติ เนื่องจากขนาดต้นไม้ที่ใหญ่ขึ้นจะใช้ทรัพยากรของเครื่องคอมพิวเตอร์ และยังเพิ่มความซับซ้อนของโครงสร้างต้นไม้ ซึ่งจะทำให้เวลาในการรันโปรแกรมนานขึ้นอีกด้วย

7. รายการอ้างอิง

- [1] ชัยวัฒน์ เจษฎาปกรณ์, การลดทอนความฟิตรพยายามเชิงคำนวณของวิธีการเรียนรู้ แบบกำหนดการเชิงพันธุกรรม, วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2540
- [2] Goldberg, D. E. and O'Reilly, U.M. *Where does the good stuff go, and why? How contextual semantics influence expression structure in simple genetic programming*, Proceedings of the First European Workshop on Genetic Programming, LNCS, Paris, 1998

[3] Rosca, J. and Ballard, D. H. *Causality in genetic programming*, Genetic Algorithms: Proceedings of the Sixth International Conference (ICGA95), pp. 256-263. Pittsburgh, PA, USA. Morgan Kaufmann, 1996

[4] Rosca, J. *Generality versus size in genetic programming*, Genetic Programming 1996: Proceedings of the First Annual Conference, pp. 381-387, Stanford University, CA, USA. Massachusetts: MIT Press, 1995

[5] Koza, J. *Genetic Programming II: Automatic Discovery of Reusable Programs*, MIT Press, 1994.