

การเรียนรู้แบบเสริมกำลังโดยใช้โมเดลสำหรับควบคุมการเคลื่อนที่ ในแนวระนาบของแขนหุ่นยนต์

MODEL-BASED REINFORCEMENT LEARNING FOR PLANAR MOTION CONTROL OF ROBOTIC ARMS

เมธา เมืองประเสริฐ¹ พิศักดิ์ เจิมประยงค์² และ กิตติพงษ์ บุญโล่ง^{3*}

¹นิสิต, สาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพา

169 ถ.ลงหาดบางแสน ต.แสนสุข อ.เมืองชลบุรี จ.ชลบุรี 20131, metamps14@gmail.com

^{2,3}อาจารย์, สาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพา

169 ถ.ลงหาดบางแสน ต.แสนสุข อ.เมืองชลบุรี จ.ชลบุรี 20131,

²pisak.ch@eng.buu.ac.th, ³kittipong@eng.buu.ac.th

Meta Mueangprasert¹, Pisak Chermprayong² and Kittipong Boonlong^{3*}

¹Student, Department of Mechanical Engineering, Faculty of Engineering Burapha University,
169 Long-Had Bangsaen Rd., Saen Suk Sub-district, Muang District, Chonburi 20131,
Thailand, metamps14@gmail.com

^{2,3}Lecturer, Department of Mechanical Engineering, Faculty of Engineering Burapha University,
169 Long-Had Bangsaen Rd., Saen Suk Sub-district, Muang District, Chonburi 20131, Thailand,
²pisak.ch@eng.buu.ac.th, ³kittipong@eng.buu.ac.th

บทคัดย่อ

งานวิจัยนี้เป็นการนำเสนอวิธีการเรียนรู้แบบเสริมกำลังโดยใช้โมเดล (MBRL) สำหรับควบคุมการเคลื่อนที่ในแนวระนาบของแขนหุ่นยนต์ 2 องศาอิสระและ 3 องศาอิสระ โดยทดสอบผ่านกรณีศึกษา 3 ปัญหา คือ ปัญหาการวางวัตถุ ปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระและ 3 องศาอิสระที่มีการเพิ่มสัญญาณรบกวนการเคลื่อนที่และมีเงื่อนไขการฝึกสอนที่แตกต่างกัน พร้อมทั้งเปรียบเทียบวิธีจลนศาสตร์แบบผกผัน (IK) กับวิธี MBRL ที่ใช้เทคนิคการถดถอยการเรียนรู้ของเครื่อง 3 เทคนิค คือ การถดถอยของกระบวนการเกาส์เซียน (GPR) โครงข่ายประสาทเทียม (ANN) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (SVR) ในการสร้างแบบจำลองสิ่งแวดล้อมร่วมกับวิธีการหาค่าเหมาะสมที่สุดกลยุทธวิวัฒนาการการปรับตัวของเมทริกซ์ความแปรปรวนร่วม (CMA-ES) จากผลการทดลองพบว่าการใช้เทคนิค GPR ร่วมกับวิธี CMA-ES จะให้ประสิทธิภาพสูงที่สุด เนื่องจากเทคนิค GPR เป็นการประมาณความแปรปรวนสัมพัทธ์ที่พิจารณา

ถึงสัญญาณรบกวน ดังนั้นจึงทำให้ผลอัตราความสำเร็จของเทคนิค GPR มีค่าเท่ากับ 100%, 96-100% และ 98-100% ในปัญหาการวางวัตถุ ปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระและ 3 องศาอิสระตามลำดับ ซึ่งสูงกว่าเทคนิค ANN, SVR, และ IK อย่างชัดเจน แม้ว่าเทคนิค GPR จะใช้เวลาในการฝึกสอนมากที่สุด แต่ก็ถือว่ามีความเหมาะสมกว่าเทคนิคอื่น ซึ่งมีอัตราความสำเร็จโดยเฉลี่ยประมาณเพียง 50%

คำสำคัญ: การควบคุมแขนหุ่นยนต์, การเรียนรู้แบบเสริมกำลังโดยใช้โมเดล, การถดถอยการเรียนรู้ของเครื่อง, กลยุทธ์วิวัฒนาการการปรับตัวของเมทริกซ์ความแปรปรวนร่วม

ABSTRACT

This research proposes model-based reinforcement learning (MBRL) for planar motion control of 2-DOF and 3-DOF robotic arms. Three case studies - placing task, 2-DOF and 3-DOF reaching tasks - are used as test problems. The 2-DOF and 3-DOF reaching tasks were investigated with additional noise in motion control signal and different training techniques. Within MBRL, 3 machine learning regression techniques, Gaussian process regression (GPR), artificial neural network (ANN) and support vector regression (SVR) were used to create environment model and then combined with an optimization algorithm, covariance matrix adaptation evolution strategy (CMA-ES). They were also benchmarked with the standard technique inverse kinematics (IK). The results show that MBRL with GPR and CMA-ES has the highest performance against the other 3 techniques. Since GPR is approximating covariance that considered noise, therefore, its success rates, which are 100%, 96-100% and 98-100% success rate in placing task, 2-DOF and 3-DOF reaching task respectively, was higher than those of ANN, SVR and IK, obviously. Although GPR spent the most training time, GPR was more suitable than other techniques of which the approximately average success rate was only 50%.

KEYWORDS: robotic arm control, model-based reinforcement learning, machine learning regression, covariance matrix adaptation evolution strategy

1. บทนำ

ในอดีตแขนหุ่นยนต์ได้ถูกนำไปประยุกต์ใช้งานเป็นอย่างมากในด้านอุตสาหกรรม ไม่ว่าจะเป็นการนำแขนหุ่นยนต์ไปใช้ในการเคลื่อนย้ายวัตถุ การนำไปใช้งานในพื้นที่แคบ ๆ หรือสภาพแวดล้อมที่มนุษย์ไม่สามารถเข้าไปทำงานได้ แต่ในปัจจุบันด้วยความก้าวหน้าของเทคโนโลยีจึงมีการนำ

หุ่นยนต์มาใช้งานในด้านอื่น ๆ มากยิ่งขึ้น เช่น ในด้านการแพทย์ที่นำแขนหุ่นยนต์มาช่วยในการผ่าตัดเปลี่ยนข้อเข่าเทียม [1] และช่วยดูแลผู้ป่วยโรคกล้ามเนื้ออ่อนแรง [2] ในด้านการเกษตรที่นำแขนหุ่นยนต์มาใช้สำหรับการเก็บเกี่ยวมะเขือเทศ [3]

การควบคุมการเคลื่อนที่ของแขนหุ่นยนต์เป็นสิ่งสำคัญในการใช้งานหุ่นยนต์ โดยวิธีการที่นิยมนำมาใช้ควบคุมหุ่นยนต์อย่างอัตโนมัติ คือ การเขียนโปรแกรมพีแอลซี (Programmable logic controller; PLC) ซึ่งเหมาะกับงานที่มีรูปแบบซ้ำ ๆ เป็นเวลานาน แต่อย่างไรก็ตามหากต้องการนำไปประยุกต์ใช้สำหรับงานที่ไม่ได้มีการทำงานซ้ำ ๆ หรือมีการหยิบและวางวัตถุ (Pick and Place) ที่ตำแหน่งเริ่มต้นและเป้าหมายเปลี่ยนแปลงอยู่ตลอดเวลา เช่น การช่วยเหลือผู้ป่วยที่เป็นอัมพาตในการหยิบสิ่งของต่าง ๆ หรือการหยิบสินค้าบนชั้นวางตามรายการสั่งซื้อของลูกค้าในร้านสะดวกซื้อ ก็จะต้องมีการตั้งโปรแกรมการทำงานเป็นจำนวนมาก ซึ่งอาจทำให้พื้นที่ของหน่วยความจำไม่สามารถรองรับได้ วิธีการที่ถูกนำเสนอในการแก้ไขปัญหา คือ การเรียนรู้ของเครื่อง (Machine learning) โดย Jiang et al [4] ได้นำเสนอวิธีการเรียนรู้แบบเสริมกำลังเชิงลึกที่มีพื้นฐานมาจากอัลกอริทึม DDPG แบบอสมมาตร ในการควบคุมหุ่นยนต์โดยมุ่งไปยังการทำรูปภาพของเป้าหมายให้มีประสิทธิภาพ Zhang et al [5] ได้นำเสนอวิธีการเรียนรู้แบบเสริมกำลัง PPO ที่ถูกปรับปรุงสำหรับการควบคุมแขนหุ่นยนต์ 6 องศาอิสระ อีกทั้ง Joshi et al [6] ยังได้นำเสนอวิธีการเรียนรู้แบบเสริมกำลังเชิงลึก Double deep Q-learning (DDQN) ร่วมกับ Grasp-Q-Network สำหรับการจับวัตถุของแขนหุ่นยนต์ Baxter แต่อย่างไรก็ตามวิธีการดังกล่าวจะเป็นการเรียนรู้แบบเสริมกำลังโดยไม่ใช้โมเดล ซึ่งเป็นหนึ่งในประเภทของวิธีการเรียนรู้แบบเสริมกำลังที่มีข้อเสีย คือ ต้องใช้จำนวนข้อมูลและเวลาในการฝึกสอนเป็นจำนวนมากในการแก้ไขปัญหา ในขณะที่วิธีการอีกประเภทหนึ่งสามารถแก้ไขปัญหาเหล่านี้ได้ คือ การเรียนรู้แบบเสริมกำลังโดยใช้โมเดล โดยมีงานวิจัยของ Li et al [7] ที่นำเสนอวิธีการเรียนรู้แบบเสริมกำลังโดยใช้โมเดลในการควบคุมหุ่นยนต์ Baxter และ Deisenroth et al [8] ที่นำเสนอวิธีการเรียนรู้แบบเสริมกำลังโดยใช้โมเดลสำหรับควบคุมแขนหุ่นยนต์ในปัญหาการวางซ้อนกล่อง (Block stacking task) ซึ่งวิธีการดังกล่าวสามารถลดจำนวนข้อมูลและเวลาในการฝึกสอนของกรณีศึกษาได้ อีกทั้งยังให้ประสิทธิภาพการทำงานที่ดี ดังนั้นงานวิจัยนี้จึงได้นำเสนอวิธีการเรียนรู้แบบเสริมกำลังโดยใช้โมเดลสำหรับควบคุมการเคลื่อนที่ของแขนหุ่นยนต์ 2 องศาอิสระและ 3 องศาอิสระในแนวระนาบผ่านกรณีศึกษา พร้อมทั้งเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องที่ใช้ในวิธีการที่นำเสนอร่วมกับวิธีการหาค่าเหมาะสมที่สุด เพื่อหาเทคนิคที่มีประสิทธิภาพมากที่สุดในการควบคุมการเคลื่อนที่ของแขนหุ่นยนต์

2. การเรียนรู้แบบเสริมกำลังโดยใช้โมเดล (Model-based reinforcement learning; MBRL)

การเรียนรู้แบบเสริมกำลังโดยใช้โมเดลคือหนึ่งในประเภทของการเรียนรู้แบบเสริมกำลังที่แบ่งตามการจำลองสิ่งแวดล้อมโดยการให้ตัวแทน (Agent) ได้เรียนรู้และทำความเข้าใจถึงความสัมพันธ์ของสถานะ (State) การกระทำ (Action) และรางวัล (Reward) ในสภาพแวดล้อม (Environment) ที่จำลองขึ้น เพื่อที่จะสามารถวางแผนการตัดสินใจได้ว่าควรกระทำสิ่งใดในสถานะต่าง ๆ เพื่อให้ได้รางวัลรวม (Cumulative reward) ในระยะยาวมากที่สุด [9]

2.1 แบบจำลองสิ่งแวดล้อม (Environment model)

ในการจำลองสิ่งแวดล้อมของวิธีการ MBRL เป็นการประมาณความสัมพันธ์ระหว่างค่าตัวแปรนำเข้า (Input variable; IV) และค่าตัวแปรส่งออก (Output variable; OV) โดยใช้เทคนิคการเรียนรู้ของเครื่องที่มีความเหมาะสมสำหรับปัญหาการถดถอย เพื่อวางแผนการตัดสินใจในการเลือกการกระทำต่อสถานะต่าง ๆ ซึ่งงานวิจัยนี้ได้นำมาประยุกต์ใช้งานทั้งหมด 3 เทคนิคดังนี้

1) การถดถอยของกระบวนการเกาส์เซียน (Gaussian process regression; GPR) เป็นระเบียบวิธีแบบเบย์เซียน (Bayesian) ที่ประมาณความแปรปรวนระหว่างข้อมูลฝึกสอนด้วยฟังก์ชันความแปรปรวนเฉพาะ [10] โดยฟังก์ชันที่นิยมใช้งานคือ Squared exponential function แสดงดังสมการที่ (1)

$$k(x, x') = \sigma_f^2 \exp \left[\frac{-(x - x')^2}{2l^2} \right] \quad (1)$$

โดย $k(x, x')$ คือ ฟังก์ชันความแปรปรวนเฉพาะระหว่าง x กับ x'

x และ x' คือ คุณลักษณะของข้อมูล

σ_f^2 คือ ค่าความแปรปรวน

l คือ ค่าความยาวสเกล (Length scale)

ในทางปฏิบัติแล้วข้อมูลมักจะมีสัญญาณรบกวน (Noise) ซึ่งอาจเกิดจากข้อผิดพลาด (Error) ในการวัดหรือสิ่งแวดล้อม เป็นต้น ดังนั้นจึงสามารถพิจารณาความสัมพันธ์ของข้อมูลดังสมการที่ 2

$$y = f(x) + N(0, \sigma_n^2) \quad (2)$$

โดย y คือ ค่าที่ต้องการทำนาย

$f(x)$ คือ ฟังก์ชันแฝง (Latent function)

N คือ สัญญาณรบกวนแบบเกาส์เซียน

σ_n^2 คือ ค่าความแปรปรวนในสัญญาณรบกวนแบบเกาส์เซียน

เมื่อสัญญาณรบกวนถูกพิจารณา ฟังก์ชันความแปรปรวนเฉพาะระหว่าง x กับ x' ในสมการที่

(1) จึงเปลี่ยนเป็นดังสมการที่ (3)

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^2}{2l^2}\right] + \sigma_n^2 \delta(x, x') \quad (3)$$

โดย $\delta(x, x')$ คือ Kronecker delta function ซึ่งจะเท่ากับ 1 หาก x, x' มีค่าเท่ากัน และเท่ากับ 0 หากเงื่อนไขเป็นอย่างอื่น

จากสมการที่ (3) ตัวแปร l, σ_f, σ_n จะต้องหาค่าที่เหมาะสมที่สุด โดยกำหนดให้เวกเตอร์ตัวแปรตัดสินใจ $\{\theta\} = \{l, \sigma_f, \sigma_n\}$

สมมติให้ $\{y\} = \{y_1, y_2, \dots, y_n\}$ คือ ชุดข้อมูลเอาต์พุตจำนวน n ตัว โดย x_1, x_2, \dots, x_n คือตัวแปรอินพุตของ $\{y\}$ และ y^* คือค่าเอาต์พุตที่ต้องการทำนายจากตัวแปรอินพุต x^* ในตอนเริ่มต้นต้องคำนวณฟังก์ชันความแปรปรวนดังสมการที่ (3) ของชุดข้อมูลทั้งหมดที่เป็นไปได้ ซึ่งจะได้เมทริกซ์ดังสมการที่ (4) และ (5) ดังนี้

$$[K] = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix} \quad (4)$$

$$[K_*] = [k(x_*, x_1) \quad k(x_*, x_2) \quad \dots \quad k(x_*, x_n)], \quad [K_{**}] = [k(x_*, x_*)] \quad (5)$$

โดยการจำลองแบบกระบวนการเกาส์เซียน ข้อมูลที่มีจะสามารถแทนด้วยตัวอย่างที่สร้างจากการแจกแจงแบบเกาส์เซียนหลายตัวแปร (Gaussian normal distribution) ดังนี้

$$\begin{bmatrix} \{y\} \\ y_* \end{bmatrix} \approx N\left(\{0\}, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \quad (6)$$

ทั้งนี้จะมุ่งความสนใจไปที่ความน่าจะเป็นเงื่อนไข (Conditional probability) $p(y^*|y)$ ซึ่งมีความหมายว่า “เมื่อมีชุดข้อมูล $\{y\}$ ความเป็นไปได้ที่ทำนายได้ค่าเป็น y^* เป็นเท่าใด” โดยความน่าจะเป็นนี้จะขึ้นอยู่กับตามการแจกแจงแบบเกาส์เซียนดังสมการที่ (7)

$$y^*|y \approx N\left([K]^* [K]^{-1} y, [K^{**}] - [K^*] [K]^{-1} [K^*]^T\right) \quad (7)$$

ค่าที่ดีที่สุดของการประมาณ y^* คือค่าเฉลี่ยของการแจกแจงดังกล่าว จะได้ดังสมการที่ (8)

$$\bar{y}^* = [K]^* [K]^{-1} y \quad (8)$$

ขณะที่ความไม่แน่นอนของการประมาณอธิบายได้ด้วยค่าความแปรปรวน ดังสมการที่ (9)

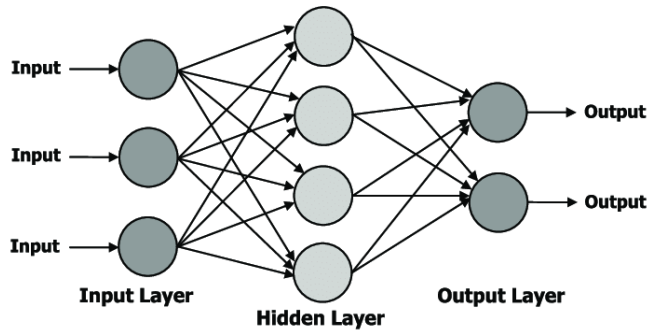
$$\text{var}(y^*) = [K^{**}] - [K^*] [K]^{-1} [K^*] \quad (9)$$

ทั้งนี้จะต้องหาค่าตัวแปรตัดสินใจ $\{\theta\} = \{\mu, \sigma_r, \sigma_n\}$ ที่เหมาะสมที่สุด โดยมีฟังก์ชันวัตถุประสงค์ที่กำหนดตามทฤษฎีของเบย์ (Bayes' theorem) ซึ่งแสดงด้วยค่าลอการิทึมความคล้าย (Likelihood Logarithm) ดังสมการที่ (10)

$$\log p(\{y\}|\{x\}, \{\theta\}) = -\frac{1}{2} \{y\}^T [K]^{-1} \{y\} - \frac{1}{2} \log(\det([K])) - \frac{n}{2} \log 2\pi \quad (10)$$

หลังจากได้ค่าตัวแปรตัดสินใจที่เหมาะสมก็ถือเป็นอันเสร็จสิ้นการจำลองของวิธี GPR

2) โครงข่ายประสาทเทียม (Artificial neural network; ANN) เป็นแบบจำลองทางคณิตศาสตร์ที่ลอกเลียนแบบการทำงานของระบบประสาทของมนุษย์ให้มาอยู่บนเครื่องคอมพิวเตอร์ เพื่อให้คอมพิวเตอร์สามารถคิด วิเคราะห์ และประมวลผลได้ในลักษณะเดียวกับโครงข่ายประสาทของมนุษย์ [11] โดยจะมีชั้นการทำงานทั้งหมด 3 ชั้น (รูปที่ 1) คือ 1) ชั้นข้อมูลนำเข้า (Input layer) เป็นชั้นที่ทำหน้าที่ในการรับข้อมูลเข้ามาในโครงข่ายประสาท และส่งต่อไปยังชั้นแอบแฝง 2) ชั้นแอบแฝง (Hidden layer) เป็นชั้นที่ทำหน้าที่ประมวลผลข้อมูลเพื่อส่งต่อไปยังชั้นข้อมูลส่งออก โดยในชั้นนี้สามารถมีได้มากกว่า 1 ชั้น และในแต่ละชั้นสามารถมีได้มากกว่า 1 โหนด 3) ชั้นข้อมูลส่งออก (Output layer) เป็นชั้นที่ทำหน้าที่รับข้อมูลจากชั้นแอบแฝงในชั้นสุดท้าย และทำการแสดงผลที่ได้จากการประมวลผลในโครงข่ายประสาทเทียม



รูปที่ 1 ชั้นการทำงานของโครงข่ายประสาทเทียม [12]

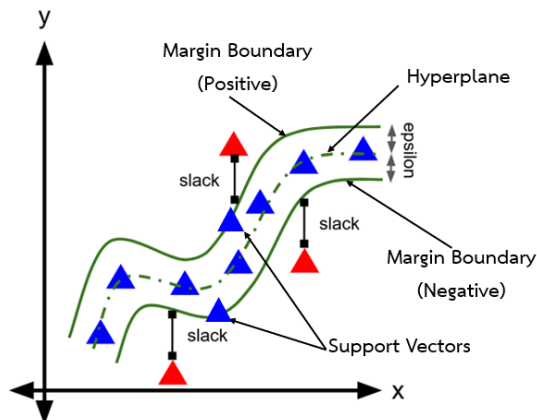
3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support vector regression; SVR) เป็น การสร้างแนวแบ่งข้อมูล (Hyperplane) ด้วยฟังก์ชันเคอร์เนล (Kernel function) เพื่อใช้ในการ ทำนายค่า (สมการที่ 11 แสดงฟังก์ชันเคอร์เนลเรเดียลเบสิส (Radial basis function; RBF)) โดย ทหารยะห่างระหว่างเส้นขอบเขต (Margin boundary) กับแนวแบ่งข้อมูลที่มีค่าเหมาะสมที่สุด ซึ่งจุด ที่อยู่บนเส้นขอบเขตจะเรียกว่า ซัพพอร์ตเวกเตอร์ (Support vectors) [13] แสดงดังรูปที่ 2

$$K(x, x') = \exp \left[-\frac{\|x - x'\|^2}{2\sigma^2} \right] \quad (11)$$

โดย $K(x, x')$ คือ ฟังก์ชันความแปรปรวนระหว่าง x กับ x'

x และ x' คือ คุณลักษณะของข้อมูล

σ คือ ค่าเบี่ยงเบนมาตรฐาน

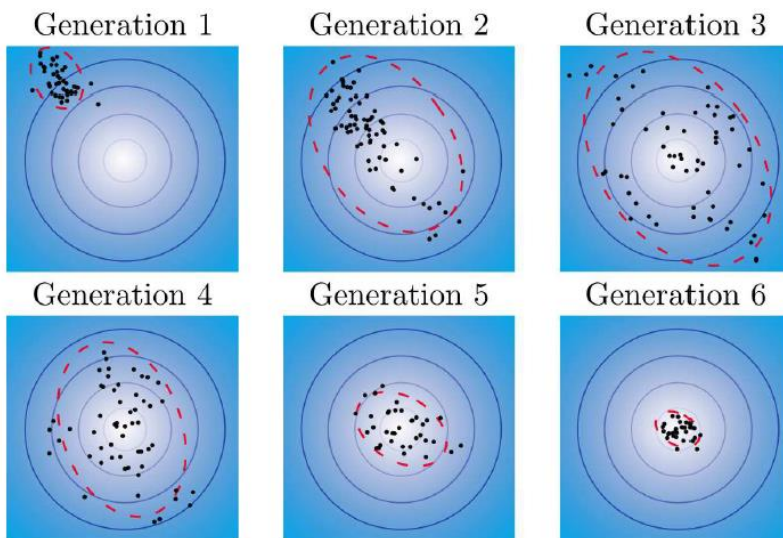


รูปที่ 2 องค์ประกอบของซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย [14]

2.2 วิธีการหาค่าเหมาะสมที่สุด (Optimization algorithm)

ในการเลือกการกระทำต่อสถานะต่าง ๆ จะประยุกต์ใช้วิธีการหาค่าเหมาะสมที่สุดเพื่อหาการกระทำที่ดีที่สุดจากแบบจำลองสิ่งแวดล้อมที่สร้างขึ้น ซึ่งในงานวิจัยนี้ได้เลือกใช้วิธีกลยุทธ์วิวัฒนาการการปรับตัวของเมทริกซ์ความแปรปรวนร่วม (Covariance matrix adaptation evolution strategy; CMA-ES) ที่เป็นการใช้เทคนิคของเมทริกซ์ความแปรปรวนร่วมในการหาค่าตอบและเหมาะสมสำหรับปัญหาที่มีความสัมพันธ์แบบไม่เชิงเส้น อีกทั้งยังเป็นวิธีการหาค่าเหมาะสมที่สุดแบบสุ่มโดยไม่ใช้อนุพันธ์ (Derivative-free optimization) [15]

ขั้นตอนในการหาค่าตอบของ CMA-ES (รูปที่ 3) โดยสรุปเป็นขั้นตอนดังแสดงเป็นแต่ละข้อ โดยตัวแปรที่เขียนอยู่ในเครื่องหมาย {...} แทนเวกเตอร์ที่มีขนาดเท่ากับ $n \times 1$ โดย (n) เป็นจำนวนตัวแปรที่ทำการหาค่าเหมาะสมที่สุด ขณะที่ตัวแปรที่อยู่ในเครื่องหมาย [...] คือเมทริกซ์ขนาด $n \times n$ และตัวแปรที่อยู่นอกเหนือจากที่กล่าวมาเป็นค่าสเกลาร์



รูปที่ 3 การหาค่าตอบของ CMA-ES [16]

(1) กำหนดตัวแปรค่าคงที่ได้แก่ ขนาดประชากรคำตอบ (Population size; λ) ขนาดตัวอย่าง (Sample size) จำนวนคำตอบรุ่นลูก (Number of offspring) และ μ เป็นจำนวนคำตอบรุ่นพ่อแม่ (Parent number) ที่ถูกเลือก จำนวนจุดค้นหาที่ถูกเลือกในประชากร (Number of selected search points in the population) โดยที่ $\mu \leq \lambda$

(2) กำหนดค่าเริ่มต้นของค่าเฉลี่ย (Mean; m) ขนาดก้าว (Step size; σ) เมทริกซ์ความแปรปรวน $[C] =$ เมทริกซ์เอกลักษณ์ (Identity matrix) $[I]$ เส้นทางวิวัฒนาการแบบไอโซทรอปิก

(Isotropic evolution path) $\{p_s\} = \{0\}$ เส้นทางวิวัฒนาการแบบแอนไอโซทรอปิก (Anisotropic evolution path) $\{p_s\} = \{0\}$

(3) สุ่มค่าตอบใหม่แล้วประเมินสร้างคำตอบ x ใหม่จากการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal distribution) ที่มีตำแหน่งค่าเฉลี่ย $= \{m\}$ และเมทริกซ์ความแปรปรวน $[V] = \sigma^2[C]$

(4) คำนวณค่าความแข็งแรง (Fitness; f_i) ของ $\{x_i\}$ แต่ละตัวจากฟังก์ชันที่ต้องการหาค่าเหมาะสมที่สุด

(5) จัดเรียง $\{x_1\}, \{x_2\}, \dots, \{x_n\}$ ตามค่าความแข็งแรง โดย $\{x_1\}$ จะให้ค่าความแข็งแรงมากที่สุดดังนั้นจึงเป็นคำตอบที่ดีที่สุด

(6) อัปเดต $\{m\}$ โดยใช้ค่า $\{x_1\}, \{x_2\}, \dots, \{x_n\}$

(7) อัปเดต $\{p_s\}$ โดยใช้ค่าปัจจุบันของ $\{p_s\}, \sigma, [C]$ และ $\{m\}-\{m'\}$

(8) อัปเดต $\{p_s\}$ โดยใช้ค่าปัจจุบันของ $\{p_s\}, \sigma, \{m\}-\{m'\}$ และ $\{p_s\}$

(9) อัปเดต $[C]$ โดยใช้ค่าปัจจุบันของ $[C], \{p_s\}, \{x\}-\{m'\}, \sigma$

(10) อัปเดต σ โดยใช้ค่าปัจจุบันของ σ และ $\{p_s\}$

(11) สิ้นสุดการค้นหาคำตอบ เมื่อจำนวนรอบซึ่งนับจากขั้นตอนที่ 3 - 10 ครบตามค่าที่กำหนดไว้ และจะได้ $\{x_1\}$ เป็นผลลัพธ์ของการค้นหาคำตอบเหมาะสมที่สุดของ CMA-ES

3. การดำเนินการของวิธีการเรียนรู้แบบเสริมกำลังโดยใช้โมเดล

ในการฝึกสอน (Training) ของวิธี MBRL จะเริ่มต้นจากการให้ตัวแทนเคลื่อนที่แบบสุ่มในแต่ละครั้ง (Step) ของรอบการทำงาน (Episode) พร้อมเก็บข้อมูลของสถานะ (State; S) การกระทำ (Action; A) และสถานะถัดไป (Next state; S') จนครบจำนวนรอบที่กำหนดไว้เพื่อนำมาใช้ในการสร้างแบบจำลองสิ่งแวดล้อมด้วยเทคนิคการเรียนรู้ของเครื่อง โดยมีข้อมูลนำเข้า คือ สถานะและการกระทำ ข้อมูลส่งออก คือ สถานะถัดไป จากนั้นในรอบการทำงานถัดไปจึงกำหนดตำแหน่งเป้าหมายที่ต้องการให้ตัวแทนไปถึงแบบสุ่มที่มีการกระจายตัวอย่างสม่ำเสมอ และนำขั้นตอนวิธีการหาค่าเหมาะสมที่สุดมาใช้ในการหาการกระทำที่ดีที่สุดของแต่ละครั้งการเคลื่อนที่จากแบบจำลองสิ่งแวดล้อมที่สร้างขึ้น พร้อมเก็บข้อมูลของสถานะ การกระทำและสถานะถัดไปจนครบจำนวนครั้งสูงสุดที่กำหนดไว้หรือเคลื่อนที่ไปถึงเป้าหมายในแต่ละรอบ แล้วนำข้อมูลที่ได้ไปรวมกับชุดข้อมูลเดิมเพื่อสร้างแบบจำลองสิ่งแวดล้อมอีกครั้ง และทำซ้ำจนกระทั่งตรงตามเงื่อนไขของการหยุดสร้างแบบจำลองสิ่งแวดล้อม เช่น ครบจำนวนรอบสูงสุดที่กำหนดไว้ในการสร้างแบบจำลองสิ่งแวดล้อมหรือตัวแทนเคลื่อนที่ไปถึงเป้าหมายติดต่อกันเป็นจำนวน N รอบ จากนั้นจึงนำแบบจำลองสิ่งแวดล้อมมาใช้ในการทดสอบ (Test) เพื่อหาประสิทธิภาพของแบบจำลอง

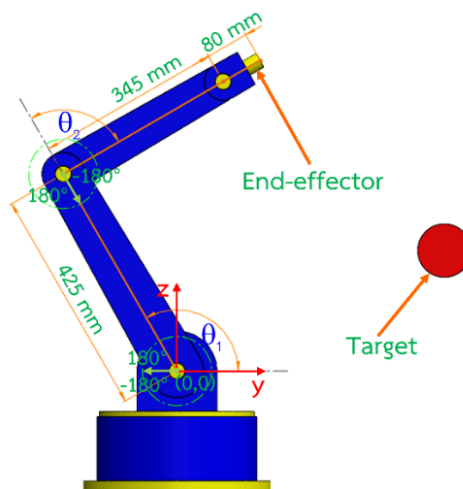
4. กรณีศึกษา

4.1 ปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระ (2-DOF Reaching task)

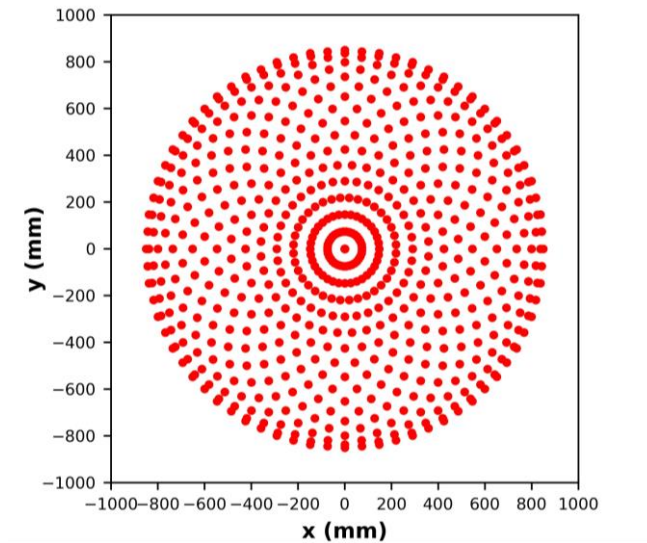
ปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระเป็นการควบคุมการเคลื่อนที่แบบเชิงมุม (Angular motion) ด้วยการหมุนของมอเตอร์ข้อต่อ 2 มุมในระนาบ yz เพื่อให้ปลายแขนหุ่นยนต์ (End-effector) ไปถึงตำแหน่งเป้าหมายที่ต้องการ 0 ซึ่งแบ่งการทดสอบออกเป็น 2 กรณี คือ

1) การควบคุมการหมุนของมอเตอร์ข้อต่อที่มีการกำหนดจำนวนรอบสูงสุดในการสร้างแบบจำลองสิ่งแวดล้อมและการเพิ่มสัญญาณรบกวนแบบเอกรูป (Uniform noise) ของการหมุนที่มีค่าเท่ากับ 0.0% 2.5% 5.0% 7.5% 10.0% 15.0% และ 20.0% เนื่องจากในทางปฏิบัติแล้วหากต้องการควบคุมให้มอเตอร์ข้อต่อหมุนไปยังมุมที่ต้องการ อาจมีปัจจัยที่ส่งผลให้มอเตอร์ที่วัดได้จริงมีความคลาดเคลื่อนเกิดขึ้นได้ เช่น สัญญาณทางไฟฟ้าที่ป้อนให้กับมอเตอร์ ข้อต่อ โดยหากส่งสัญญาณการควบคุมให้มอเตอร์ข้อต่อหมุนไปยังมุม 40 องศา มุมที่เป็นไปได้จากการหมุนจะมีค่าอยู่ระหว่าง 36 ถึง 44 องศาในกรณีที่มีสัญญาณรบกวนเท่ากับ 10.0%

ในแต่ละครั้งการเคลื่อนที่ มอเตอร์ข้อต่อทั้ง 2 มอเตอร์สามารถหมุนได้ระหว่าง -45 ถึง 45 องศา และมีขอบเขตของการหมุนอยู่ระหว่าง -180 ถึง 180 องศา แสดงดังรูปที่ 4 ดังนั้นจึงทำให้มีตำแหน่งเป้าหมายที่เป็นไปได้แสดงดังรูปที่ 5

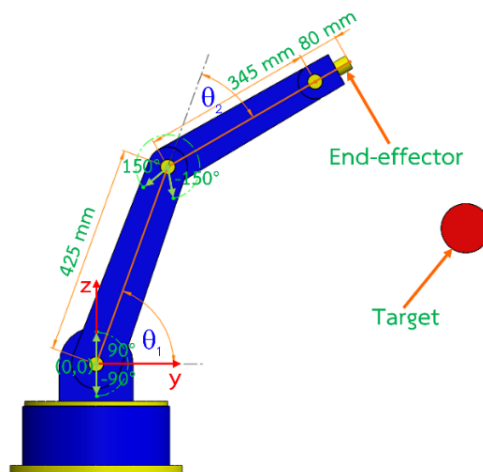


รูปที่ 4 การจำลองปัญหา 2-DOF Reaching task กรณีที่ 1

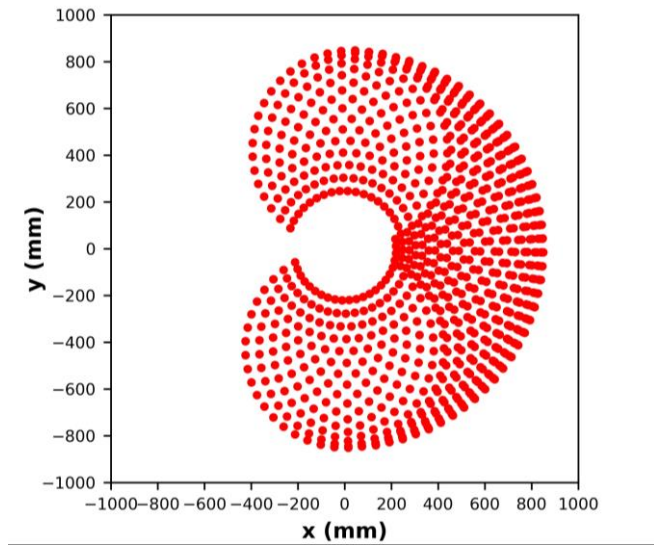


รูปที่ 5 ตำแหน่งเป้าหมายที่เป็นไปได้ของปัญหา 2-DOF Reaching task กรณีที่ 1

2) การควบคุมการหมุนของมุมข้อต่อที่ไม่มีกำหนดจำนวนรอบสูงสุดในการสร้างแบบจำลองสิ่งแวดล้อมและไม่มี的增加สัญญาณรบกวนของการหมุน โดยในแต่ละครั้งของการเคลื่อนที่ มุมข้อต่อทั้ง 2 มุมจะสามารถหมุนได้ระหว่าง -11.46 ถึง 11.46 องศา และมีขอบเขตการหมุนของมุมข้อต่อที่ 1 อยู่ระหว่าง -90 ถึง 90 องศา มุมข้อต่อที่ 2 อยู่ระหว่าง -150 ถึง 150 องศา แสดงดังรูปที่ 6 ดังนั้นจึงทำให้มีตำแหน่งเป้าหมายที่เป็นไปได้แสดงดังรูปที่ 7



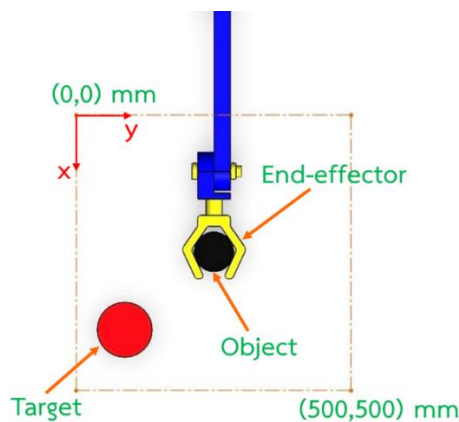
รูปที่ 6 การจำลองปัญหา 2-DOF Reaching task กรณีที่ 2



รูปที่ 7 ตำแหน่งเป้าหมายที่เป็นไปได้ของปัญหา 2-DOF Reaching task กรณีที่ 2

4.2 ปัญหาการวางวัตถุ (Placing task)

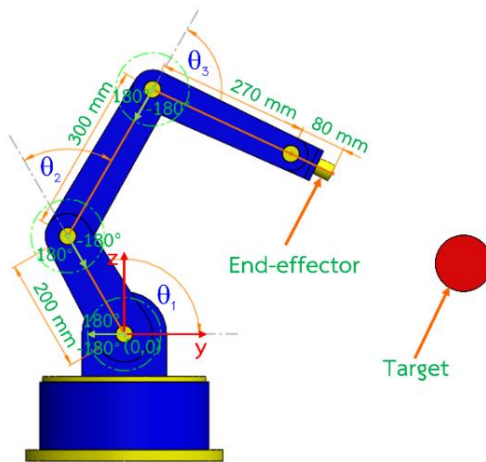
ปัญหาการวางวัตถุเป็นการนำวัตถุไปวางยังตำแหน่งเป้าหมายที่ต้องการ ผ่านการควบคุมการเคลื่อนที่แบบเชิงเส้น (Linear motion) ของปลายแขนหุ่นยนต์ในระนาบ xy [4] โดยในแต่ละครั้งของการเคลื่อนที่จะสามารถเคลื่อนที่ได้ระหว่าง -30 ถึง 30 มม. ทั้งในแนวแกน x และแกน y อีกทั้งยังมีขอบเขตของการเคลื่อนที่และตำแหน่งเป้าหมายในแต่ละแกนอยู่ภายใน 500 มม. แสดงดังรูปที่ 8



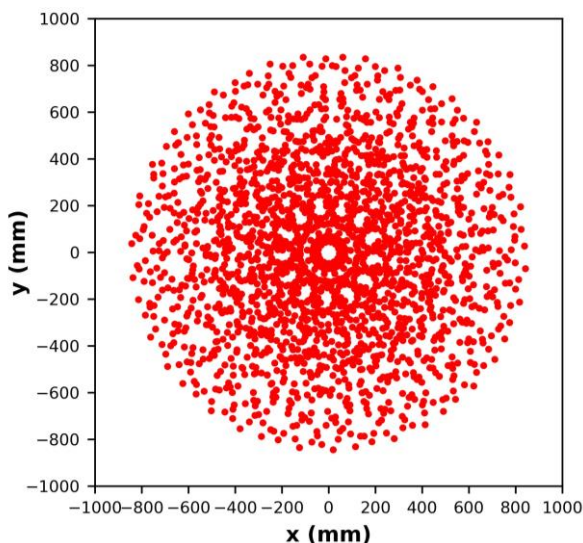
รูปที่ 8 การจำลองปัญหาการวางวัตถุ

4.3 ปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 3 องศาอิสระ (3-DOF Reaching task)

ปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 3 องศาอิสระจะมีวัตถุประสงค์ของปัญหาเหมือนกับปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระ แต่จะเป็นการควบคุมการหมุนของมอเตอร์ 3 มุม โดยในแต่ละครั้งการเคลื่อนที่ มุมข้อต่อทั้ง 3 มุมจะสามารถเคลื่อนที่ด้วยการหมุนระหว่าง -45 ถึง 45 องศา และมีขอบเขตการหมุนอยู่ระหว่าง -180 ถึง 180 องศา แสดงดังรูปที่ 9 ในทำนองเดียวกันกับปัญหา 2-DOF Reaching task กรณีที่ 1 ตำแหน่งเป้าหมายที่เป็นไปได้แสดงดังรูปที่ 10 อีกทั้งยังมีการเพิ่มสัญญาณรบกวนแบบเอกรูปของการหมุนที่เท่ากับ 0.0% 5.0% และ 10.0%



รูปที่ 9 การจำลองปัญหา 3-DOF Reaching task



รูปที่ 10 ตำแหน่งเป้าหมายที่เป็นไปได้ของปัญหา 3-DOF Reaching task

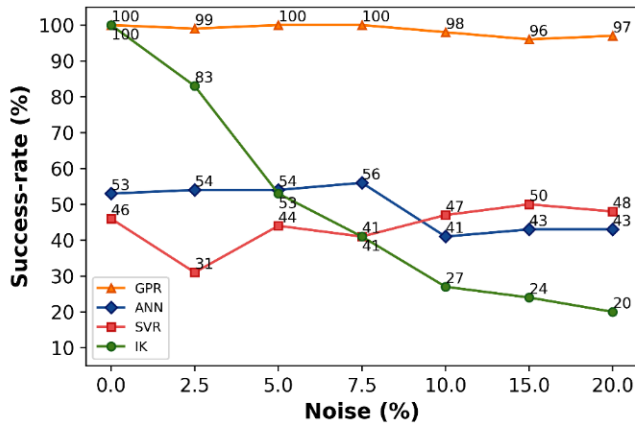
ในการกำหนดสถานะและการกระทำของแต่ละปัญหา จะกำหนดให้สถานะเป็นตำแหน่งพิกัดปลายแขนหุ่นยนต์ในระนาบ yz การกระทำเป็นการหมุนของมุมข้อต่อทั้ง 2 มุมและ 3 มุมสำหรับปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระและ 3 องศาอิสระตามลำดับ แต่ในส่วนของปัญหาการวางวัตถุ จะกำหนดให้ตำแหน่งพิกัดปลายแขนหุ่นยนต์ในระนาบ xy เป็นสถานะ การเคลื่อนที่ในแกน x และแกน y เป็นการกระทำ อีกทั้งการให้รางวัลของทุกปัญหาจะกำหนดให้มีค่าเท่ากับระยะห่างระหว่างปลายแขนหุ่นยนต์กับเป้าหมาย

5. ผลการทดลอง

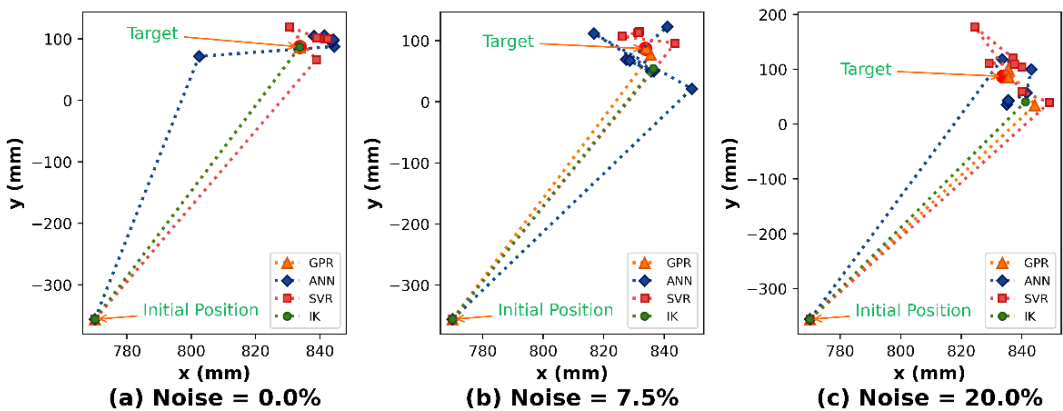
งานวิจัยนี้ดำเนินการทดสอบผ่านการเขียนโปรแกรมภาษาไพทอน (Python) ด้วยคอมพิวเตอร์ที่ใช้ GPU รุ่น NVIDIA GeForce MX150 และได้กำหนดรายละเอียดเทคนิคการเรียนรู้ของเครื่องที่นำมาใช้ในวิธีการ MBRL เพื่อเปรียบเทียบประสิทธิภาพการทำงาน โดยเทคนิค GPR จะใช้ฟังก์ชันค่าคงที่ (Constant kernel) ฟังก์ชันเรเดียลเบสิส (Radial basis function) และฟังก์ชันสีขาว (White kernel) เป็นฟังก์ชันเคอร์เนล เทคนิค ANN จะใช้จำนวนชั้นแอบแฝงเท่ากับ 6 ชั้นและในแต่ละชั้นมีจำนวน 30 โหนด อีกทั้งยังใช้ฟังก์ชันกระตุ้น คือ ฟังก์ชันเรลู (ReLU) ในชั้นแอบแฝงและฟังก์ชันไฮเปอร์โบลิคแทนเจนต์ (Tanh) ในชั้นข้อมูลส่งออก และเทคนิค SVR ใช้ฟังก์ชันเรเดียลเบสิสเป็นฟังก์ชันเคอร์เนล ซึ่งประสิทธิภาพการทำงานจะวัดด้วยอัตราความสำเร็จ (Success rate) ที่คำนวณจากการทดสอบ (Test) การเคลื่อนที่ไปยังเป้าหมายที่เกิดจากการสุ่ม โดยมีระยะห่างระหว่างปลายแขนหุ่นยนต์กับเป้าหมายที่ยอมรับได้น้อยกว่าหรือเท่ากับ 10 มม. เป็นจำนวน 100 รอบสุดท้ายในการฝึกสอน อีกทั้งการทดสอบในทุกปัญหาจะไม่พิจารณาถึงความเร็วและความแรงในการเคลื่อนที่ของแขนหุ่นยนต์เป็นสถานะในการเรียนรู้ ซึ่งเป็นทำนองเดียวกันกับงานวิจัยของ Libera et al [17] โดยในทางปฏิบัติแล้ว ส่วนใหญ่เช่นเซอร์วัดความเร็วและความแรงจะไม่ค่อยมีการนำมาใช้งานเนื่องจากมีต้นทุนที่สูง ทั้งนี้ในกรณีนี้ความเร็วและความแรงก็จะคำนวณด้วยค่าเชิงตัวเลขจากตำแหน่งของปลายแขนหุ่นยนต์ ทำให้มีความแตกต่างกันระหว่างสัญญาณจริงกับสัญญาณโดยการประมาณ ซึ่งค่าที่แตกต่างกันนี้จะส่งผลให้เกิดสัญญาณรบกวนกับตำแหน่งของปลายแขนหุ่นยนต์ รวมถึงในการควบคุมการเคลื่อนที่ของแขนหุ่นยนต์ในทางปฏิบัติ ก็อาจจะมีตำแหน่งที่คลาดเคลื่อนไปเนื่องจากแรงเฉื่อยที่เกิดขึ้นเนื่องจากความเร็วและความแรง รวมถึงความคลาดเคลื่อนเนื่องจากเซ็นเซอร์วัดตำแหน่ง ซึ่งความคลาดเคลื่อนดังกล่าวนี้ก็ถือเป็นสัญญาณรบกวนต่อตำแหน่งของปลายแขนหุ่นยนต์เช่นเดียวกัน ดังนั้นงานวิจัยนี้จึงพิจารณาทั้งกรณีศึกษาที่ไม่มีสัญญาณรบกวนและมีสัญญาณรบกวนเพื่อให้สอดคล้องกับการประยุกต์ใช้ต่อไป โดยปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณีที่มีการเพิ่มสัญญาณรบกวนได้กำหนดให้แขนหุ่นยนต์เคลื่อนที่แบบสุ่ม 1 รอบการทำงานจำนวน 7 ครั้ง และมีจำนวนรอบสูงสุดในการสร้าง

แบบจำลองสิ่งแวดล้อมเท่ากับ 250 รอบ ซึ่งในแต่ละรอบสามารถเคลื่อนที่ได้สูงสุด 7 ครั้ง อีกทั้งตำแหน่งมุมข้อต่อเริ่มต้นจะเกิดจากการสุ่มในทุกรอบการทำงาน

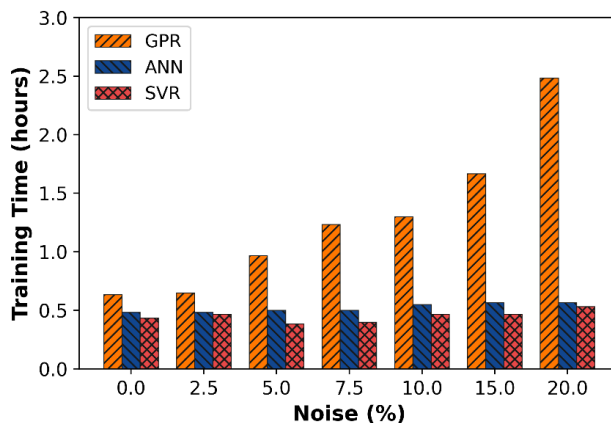
จากรูปที่ 11 เป็นการเปรียบเทียบอัตราความสำเร็จของแต่ละเทคนิคในวิธี MBRL และวิธี จลนศาสตร์แบบผกผัน (Inverse kinematic; IK) ที่เป็นการคำนวณในทางทฤษฎี ซึ่งงานวิจัยนี้ได้ใช้วิธีการหาค่าเหมาะสมที่สุดในการคำนวณหามุมของข้อต่อต่าง ๆ เมื่อทราบตำแหน่งพิกัดเป้าหมาย [18] โดยพบว่าเทคนิค GPR ให้อัตราความสำเร็จสูงที่สุดในทุกสัญญาณรบกวน เนื่องจากพื้นฐานของเทคนิค GPR จะสามารถจัดการกับสัญญาณรบกวนได้ดี แม้สัญญาณรบกวนเพิ่มขึ้น [10] แต่ในเทคนิค ANN และ SVR มีอัตราความสำเร็จใกล้เคียงกันในทุกสัญญาณรบกวนเนื่องจากการเพิ่มขึ้นของสัญญาณรบกวนไม่ได้ส่งผลถึงความแม่นยำของแบบจำลอง อีกทั้งเทคนิค GPR ยังสามารถไปถึงเป้าหมายได้สำเร็จ ด้วยจำนวนครั้งการเคลื่อนที่ที่น้อยกว่าเทคนิค ANN และ SVR ดังรูปที่ 12 ซึ่งแสดงเส้นทางการเคลื่อนที่ของแต่ละเทคนิคจากตำแหน่งเริ่มต้นที่มีพิกัดเท่ากับ (770.03, -356.15) มม. ไปยังตำแหน่งเป้าหมายที่มีพิกัดเท่ากับ (833.76, 87.00) มม. ในกรณีตัวอย่างที่มีสัญญาณรบกวนเท่ากับ 0.0% 7.5% และ 20.0% และสำหรับวิธี IK พบว่าอัตราความสำเร็จมีค่าลดลงเมื่อสัญญาณรบกวนเพิ่มขึ้น เนื่องจากการเพิ่มขึ้นของสัญญาณรบกวน ทำให้ข้อต่อมีความคลาดเคลื่อนในการหมุนมากยิ่งขึ้น จึงส่งผลให้ปลายแขนหุ่นยนต์ไปถึงตำแหน่งเป้าหมายได้ยากขึ้น ตัวอย่างเช่น หากตำแหน่งพิกัดเป้าหมายเท่ากับ (833.76, 87.00) มม. กรณีไม่มีสัญญาณรบกวนจะสามารถคำนวณหามุมข้อต่อที่ 1 และ 2 ได้เท่ากับ 15.5 และ -19.0 องศาตามลำดับ และสามารถหมุนไปยังตำแหน่งเป้าหมายได้สำเร็จ แต่หากมีสัญญาณรบกวนเท่ากับ 20% อาจจะทำให้แขนหุ่นยนต์หมุนไปยังมุม 13.7 และ -21.6 องศาที่มีพิกัดเท่ากับ (841.12, 40.62) มม. ซึ่งไม่สามารถไปถึงเป้าหมายได้ เนื่องจากมีระยะห่างระหว่างปลายแขนหุ่นยนต์กับเป้าหมายเท่ากับ 44.75 มม. โดยเส้นทางการเคลื่อนที่สำหรับวิธี IK แสดงดังรูปที่ 12 แต่อย่างไรก็ตาม เมื่อพิจารณาถึงเวลาในการฝึกสอน (Training time) ดังรูปที่ 13 จะพบว่าเทคนิค GPR ใช้เวลาการฝึกสอนมากที่สุดในทุกสัญญาณรบกวน เนื่องจากใช้เมทริกซ์ความแปรปรวนร่วมในการคำนวณเพื่อสร้างแบบจำลอง [10] หากสัญญาณรบกวนเพิ่มขึ้น ความยากในการคำนวณจึงยากขึ้นไปด้วย แต่ในเทคนิค ANN และ SVR มีเวลาในการฝึกสอนใกล้เคียงกันในทุกสัญญาณรบกวน เนื่องจากมีพื้นฐานในการคำนวณที่รวดเร็ว และสัญญาณรบกวนที่เพิ่มขึ้นไม่ได้มีผลต่อความยากในการคำนวณ



รูปที่ 11 การเปรียบเทียบอัตราความสำเร็จในแต่ละเทคนิคของปัญหา 2-DOF Reaching task กรณีที่ 1



รูปที่ 12 เส้นทางเคลื่อนที่ในแต่ละเทคนิคของปัญหา 2-DOF Reaching task กรณีที่ 1



รูปที่ 13 เวลาการฝึกสอนในแต่ละเทคนิคของปัญหา 2-DOF Reaching task กรณีที่ 1

จากผลการทดสอบของปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณี 1 พบว่าวิธี MBRL ที่ใช้เทคนิค GPR จะให้อัตราความสำเร็จสูงที่สุด แม้ว่าจะใช้เวลาฝึกสอนมากที่สุด แต่ในช่วงที่มีสัญญาณรบกวนต่ำ ๆ ก็ใช้เวลาที่ใกล้เคียงกับเทคนิค ANN และ SVR อีกทั้งในทางปฏิบัติแล้ว หากแขนหุ่นยนต์สามารถไปถึงเป้าหมายได้ ก็ถือว่าบรรลุวัตถุประสงค์ของปัญหา ซึ่งเทคนิค GPR ให้อัตราความสำเร็จโดยเฉลี่ยมากถึง 96-100% ในขณะที่เทคนิคอื่น ให้อัตราความสำเร็จโดยเฉลี่ยเพียง 50% เท่านั้น ดังนั้นจึงนำวิธี MBRL ที่ใช้เทคนิค GPR และ CMA-ES มาใช้ในการทดสอบสำหรับปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณีที่ 2 และปัญหาการวางวัตถุ หากปลายแขนหุ่นยนต์เคลื่อนที่ไปถึงเป้าหมายติดต่อกันจำนวน 100 รอบ จึงถือเป็นการสิ้นสุดการสร้างแบบจำลองสิ่งแวดล้อม จากนั้นจึงวัดอัตราความสำเร็จ ซึ่งปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณีที่ 2 กำหนดให้แขนหุ่นยนต์เคลื่อนที่แบบสุ่มจำนวน 22 ครั้งใน 1 รอบการทำงาน และสามารถเคลื่อนที่ได้สูงสุด 22 ครั้งในแต่ละรอบการฝึกสอน และตำแหน่งมุมข้อต่อเริ่มต้นในแต่ละรอบอยู่ที่ 0 องศาของทั้ง 2 มุม สำหรับปัญหาการวางวัตถุกำหนดให้ปลายแขนหุ่นยนต์เคลื่อนที่แบบสุ่มจำนวน 5 ครั้งใน 1 รอบการทำงาน และสามารถเคลื่อนที่ได้สูงสุด 17 ครั้งในแต่ละรอบการฝึกสอน โดยพิกัดปลายแขนหุ่นยนต์เริ่มต้นในแต่ละรอบอยู่ที่ (250, 250) มม.

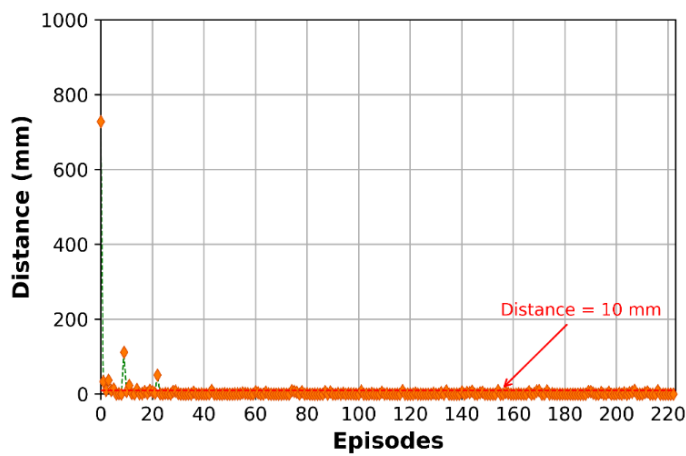
จากตารางที่ 1 และตารางที่ 2 จะพบว่าข้อมูลที่ใช้ในการฝึกสอนทั้งหมดของปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณีที่ 2 และปัญหาการวางวัตถุมีค่าน้อยกว่าการศึกษา ก่อนหน้า [4] ที่ใช้ข้อมูลในการฝึกสอนมากถึง 360,000 และ 270,000 ข้อมูลตามลำดับ อีกทั้งยังใช้เวลาในการฝึกสอนน้อยกว่าการศึกษาก่อนหน้าที่ใช้เวลามากถึง 15 ชั่วโมง 24 นาที และ 14 ชั่วโมง 1 นาทีตามลำดับ ในขณะที่อัตราความสำเร็จมีค่ามากกว่าการศึกษาก่อนหน้าที่มีอัตราความสำเร็จเพียง 95% และ 98% ตามลำดับ ซึ่งจากรูปที่ 14 ถึง รูปที่ 15 แสดงระยะห่างระหว่างปลายแขนหุ่นยนต์และเป้าหมายในแต่ละรอบการทำงานของปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณีที่ 2 และปัญหาการวางวัตถุ

ตารางที่ 1 ผลการทดสอบปัญหา 2-DOF Reaching task กรณีที่ 2

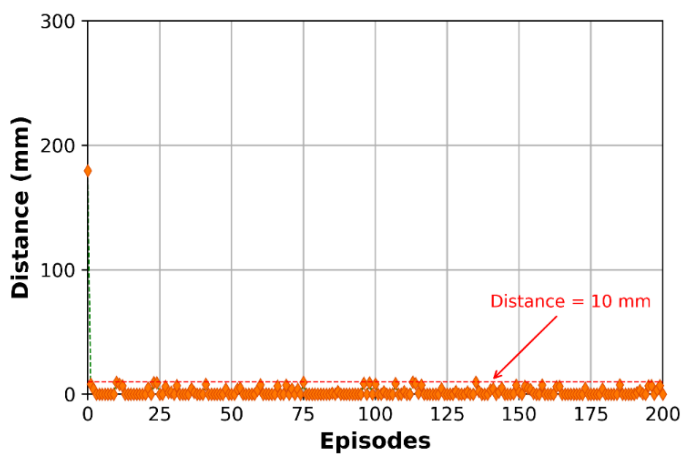
ตัวแปร	ค่า
จำนวนรอบการฝึกสอน	223
ข้อมูลในการฝึกสอน	4,906
เวลาในการฝึกสอน	3 ชั่วโมง 38 นาที
อัตราความสำเร็จ	100 %

ตารางที่ 2 ผลการทดสอบปัญหาการวางวัตถุ

ตัวแปร	ค่า
จำนวนรอบการฝึกสอน	201
ข้อมูลในการฝึกสอน	3,417
เวลาในการฝึกสอน	1 ชั่วโมง 8 นาที
อัตราความสำเร็จ	100 %



รูปที่ 14 ระยะห่างในแต่ละรอบของปัญหา 2-DOF Reaching task กรณีที่ 2

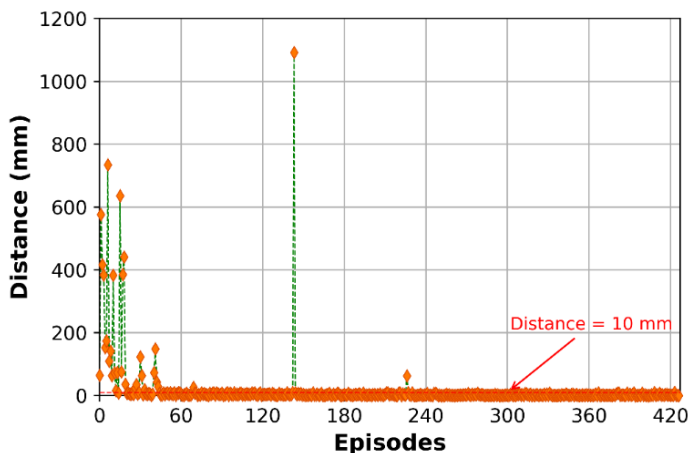


รูปที่ 15 ระยะห่างในแต่ละรอบการทำงานของปัญหาการวางวัตถุ

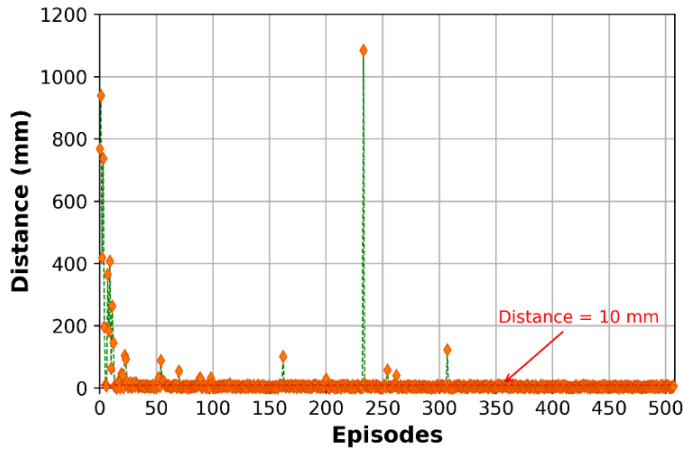
นอกจากนี้ในปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 3 องศาอิสระยังได้นำวิธี MBRL ที่ใช้เทคนิค GPR และ CMA-ES มาทำการทดสอบโดยเพิ่มสัญญาณรบกวนที่มีค่าเท่ากับ 0.0% 5.0% และ 10.0% หากปลายแขนหุ่นยนต์เคลื่อนที่ไปถึงเป้าหมายติดต่อกันจำนวน 100 รอบ จึงถือเป็นการสิ้นสุดการสร้างแบบจำลองสิ่งแวดล้อม จากนั้นจึงวัดอัตราความสำเร็จ อีกทั้งกำหนดให้แขนหุ่นยนต์เคลื่อนที่แบบสุ่มจำนวน 7 ครั้งใน 1 รอบการทำงาน และสามารถเคลื่อนที่ได้สูงสุด 7 ครั้งในแต่ละรอบการฝึกสอนโดยตำแหน่งมุมข้อต่อเริ่มต้นของทั้ง 3 มุมจะเกิดจากการสุ่มในทุกรอบ ซึ่งจากตารางที่ 3 พบว่าเมื่อสัญญาณรบกวนมีค่ามากขึ้น จำนวนรอบการฝึกสอนและเวลาในการฝึกสอนจะมีค่ามากขึ้นตามไปด้วย ในขณะที่อัตราความสำเร็จมีค่ามากถึง 98-100% โดยจากรูปที่ 16 ถึง รูปที่ 18 จะแสดงระยะห่างระหว่างปลายแขนหุ่นยนต์และเป้าหมายในแต่ละรอบการทำงานของปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 3 องศาอิสระในกรณีที่มีสัญญาณรบกวนเท่ากับ 0.0% 5.0% และ 10.0% ตามลำดับ

ตารางที่ 3 ผลการทดสอบปัญหา 3-DOF Reaching task

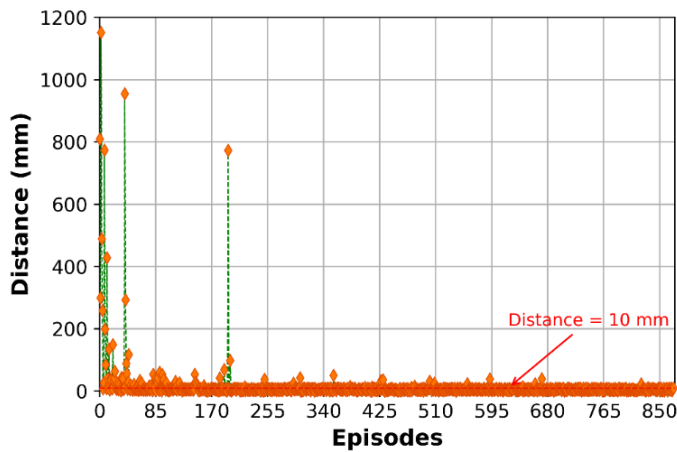
ตัวแปร	สัญญาณรบกวน		
	0.0%	5.0%	10.0%
จำนวนรอบการฝึกสอน	427	508	872
ข้อมูลในการฝึกสอน	2,989	3,556	6,104
เวลาในการฝึกสอน	1 ชั่วโมง 38 นาที	4 ชั่วโมง 58 นาที	19 ชั่วโมง 32 นาที
อัตราความสำเร็จ	100%	100%	98%



รูปที่ 16 ระยะห่างในแต่ละรอบของปัญหา 3-DOF Reaching task กรณี Noise = 0.0%



รูปที่ 17 ระยะห่างในแต่ละรอบของปัญหา 3-DOF Reaching task กรณี Noise = 5.0%



รูปที่ 18 ระยะห่างในแต่ละรอบของปัญหา 3-DOF Reaching task กรณี Noise = 10.0%

6. สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอวิธี MBRL เพื่อใช้ในการควบคุมการเคลื่อนที่ของแขนหุ่นยนต์ 2 องศาอิสระและ 3 องศาอิสระในแนวระนาบผ่านกรณีศึกษา 3 ปัญหา โดยนำเทคนิคการเรียนรู้ของเครื่อง 3 เทคนิค คือ GPR, ANN และ SVR มาใช้ในการสร้างแบบจำลองสิ่งแวดล้อมร่วมกับวิธีการหาค่าเหมาะสมที่สุด CMA-ES พร้อมทั้งเปรียบเทียบกับวิธี IK จากผลการทดสอบแสดงให้เห็นว่า การใช้เทคนิค GPR ร่วมกับวิธีการหาค่าเหมาะสมที่สุด CMA-ES จะมีประสิทธิภาพมากที่สุดเมื่อเทียบกับเทคนิคอื่น ๆ และวิธี IK โดยมีอัตราความสำเร็จสูงถึง 96-100% ในปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 2 องศาอิสระกรณีที่มีการกำหนดจำนวนรอบสูงสุดในการสร้างแบบจำลองสิ่งแวดล้อมและการเพิ่มสัญญาณรบกวนการเคลื่อนที่ อีกทั้งในปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์

2 องศาอิสระกรณีที่ไม่มีการกำหนดจำนวนรอบสูงสุดในการสร้างแบบจำลองสิ่งแวดล้อมและปัญหาการวางวัตถุ ยังให้อัตราความสำเร็จเท่ากับ 100% รวมถึงใช้ข้อมูลและเวลาในการฝึกสอนน้อยกว่าการศึกษาก่อนหน้านี้ และให้อัตราความสำเร็จสูงถึง 98-100% แม้มีการเพิ่มขึ้นของมุมข้อต่อ ในปัญหาการไปถึงเป้าหมายของแขนหุ่นยนต์ 3 องศาอิสระ ทั้งนี้เนื่องจากเทคนิค GPR เป็นระเบียบวิธีแบบเบย์เซียน ซึ่งใช้การประมาณความแปรปรวนสัมพัทธ์ที่พิจารณาถึงสัญญาณรบกวน ดังสมการที่ (3) ดังนั้นจึงทำให้ได้ผลอัตราความสำเร็จสูงกว่าเทคนิค ANN, SVR รวมถึงวิธี IK อย่างชัดเจน แม้ว่า GPR จะใช้เวลาในการฝึกสอนมากที่สุด แต่ก็ถือว่าเหมาะสมกว่าเทคนิคอื่น ซึ่งมีอัตราความสำเร็จโดยเฉลี่ยประมาณเพียง 50% เท่านั้น ทั้งนี้จากผลการทดสอบที่ได้ ยังสามารถนำวิธี MBRL ไปใช้ประโยชน์ในการฝึกสอนกับแขนหุ่นยนต์จริง โดยการใช้กล้องในการตรวจจับพิกัดตำแหน่งของปลายแขนหุ่นยนต์ ตำแหน่งวัตถุและตำแหน่งเป้าหมายเพื่อใช้เป็นข้อมูลสำหรับฝึกสอนของวิธีการ MBRL

กิตติกรรมประกาศ

ขอขอบคุณสาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพาที่ให้การสนับสนุนในการทำงานวิจัยนี้

References

- [1] Roche M. The MAKO robotic-arm knee arthroplasty system. Archives of Orthopaedic and Trauma Surgery 2021;141(12):2043-7.
- [2] Rivera GP, Eichmann C, Scherzinger S, Puck L, Roennau A, Dillmann R. Flexible, Personal service robot for ALS patients. IEEE International Conference on Robotics and Biomimetics (ROBIO); 2019. p. 1595-600.
- [3] Zhao Y, Gong L, Liu C, Huang Y. Dual-arm robot design and testing for harvesting tomato in greenhouse. IFAC-PapersOnLine 2016;49(16):161-5.
- [4] Jiang R, Wang Z, He B, Zhou Y, Li G, Zhu Z. A data-efficient goal-directed deep reinforcement learning method for robot visuomotor skill. Neurocomputing 2021;462: 389-401.
- [5] Zhang Z, Zheng C. Simulation of robotic arm grasping control based on proximal policy optimization algorithm. Journal of Physics: Conference Series 2022;2203(1):012065.

- [6] Joshi S, Kumra S, Sahin F. Robotic grasping using deep reinforcement learning. IEEE International Conference on Automation Science and Engineering (CASE); 2020. p. 1461-66.
- [7] Li X, Shang W, Cong S. Model-based reinforcement learning for robot control. IEEE International Conference on Advanced Robotics and Mechatronics (ICARM); 2020. p. 300-5.
- [8] Deisenroth MP, Rasmussen CE, Fox D. Learning to control a low-cost manipulator using data-efficient reinforcement learning. Robotics: Science and Systems VII 2011;7:57-64.
- [9] Moerland TM, Broekens J, Jonker CM. Model-based reinforcement learning: a survey. arXiv preprint arXiv:200616712 2020.
- [10] Williams C, Rasmussen C. Gaussian processes for regression. Advances in neural information processing systems 1995;8.
- [11] Abraham A. Artificial neural networks. Handbook of measuring system design 2005
- [12] Rahimi HN, Nazemizadeh M. Dynamic analysis and intelligent control techniques for flexible manipulators: a review. Advanced Robotics 2014;28(2):63-76.
- [13] Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Advances in neural information processing systems 1996;9.
- [14] Beny MA. Support vector machine: regression [Internet]. 2019 [cited 2022 Nov 15]. Available from: <https://medium.com/it-paragon/support-vector-machine-regression-cf65348b6345>
- [15] Hansen N. The CMA evolution strategy: a tutorial. arXiv preprint arXiv:160400772 2016
- [16] Tan U, Rabaste O, Adnet C, Ovarlez JP. On the eclipsing phenomenon with phase codes. IEEE International Radar Conference (RADAR); 2019. p. 1-5.
- [17] Libera DA, Romeres D, Jha DK, Yerazunis B, Nikovski D. Model-based reinforcement learning for physical systems without velocity and acceleration measurements. IEEE Robotics and Automation Letters 2020;5(2):3548-55.
- [18] Rokbani N, Alimi AM. Inverse kinematics using particle swarm optimization, a statistical analysis. Procedia Engineering 2013;64:1602-11.

ประวัติผู้เขียนบทความ



เมธา เมืองประเสริฐ นิสิตระดับมหาบัณฑิต สาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพา 169 ถ.ลงหาดบางแสน ต.แสนสุข อ.เมืองชลบุรี จ.ชลบุรี 20131

E-mail: metamps14@gmail.com

งานวิจัยที่สนใจ: Robotics, Optimization, Machine Learning, Artificial Neural Network



พิศักดิ์ เจิมประยงค์ อาจารย์ประจำสาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพา 169 ถ.ลงหาดบางแสน ต.แสนสุข อ.เมืองชลบุรี จ.ชลบุรี 20131

E-mail: pisak.ch@eng.buu.ac.th

งานวิจัยที่สนใจ: Unmanned Aerial Vehicle (UAV), Robot Kinematics and Dynamics, Robot Design



กิตติพงศ์ บุญโล่ง อาจารย์ประจำสาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพา 169 ถ.ลงหาดบางแสน ต.แสนสุข อ.เมืองชลบุรี จ.ชลบุรี 20131

E-mail: kittipong@eng.buu.ac.th

งานวิจัยที่สนใจ: Single and Multiple Objective Optimization, Evolutionary Algorithms, Mechanical Vibration, Surrogate Model Optimization, Machine Learning, Deep Learning, Reinforcement Learning

Article History:

Received: September 3, 2022

Revised: March 9, 2023

Accepted: March 9, 2023