วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

บทความวิจัย

การประยุกต์ใช้งาน Q-Learning เพื่อการจัดสรรกำลังที่เหมาะสมในระบบโนมาที่มีผู้ใช้ 2 ราย

เพชรนคร เอี่ยมสอาด ชลธิชา หวังสมัด และ กฤษฎา มามาตร*

ภาควิชาเทคโนโลยีวิศวกรรมอิเล็กทรอนิกส์, วิทยาลัยเทคโนโลยีอุตสาหกรรม,
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* ผู้ประสานงานเผยแพร่ (Corresponding Author), E-mail: kritsada.m@cit.kmutnb.ac.th

วันที่รับบทความ: 1 สิงหาคม 2565; วันที่ทบทวนบทความ: 28 พฤศจิกายน 2565; วันที่ตอบรับบทความ: 20 มกราคม 2566
วันที่เผยแพร่ออนไลน์: 13 เมษายน 2566

บทคัดย่อ: บทความนี้พิจารณาการเข้าถึงหลายส่วนแบบไม่ตั้งฉาก (Non-Orthogonal Multiple Access: NOMA) หรือ โนมา ซึ่งเป็นวิธีการถึงช่องสัญญาณของผู้ใช้ในระบบสื่อสารไร้สายยุคที่ 5 และหลังจากนั้นโดยวิธี Successive Interference Cancellation (SIC) ถูกนำมาประยุกต์ใช้เพื่อตรวจจับข้อมูลของผู้ใช้แต่ละรายในโดเมนกำลังและการจัดสรรกำลังส่งมีผลกระทบต่อสมรรถนะของระบบ บทความนี้นำเสนอการประยุกต์ใช้วิธี Q-Learning ซึ่งเป็นวิธีหนึ่งของการเรียนรู้ของเครื่องเพื่อแก้ปัญหาการจัดสรรกำลังที่เหมาะสมในระบบโนมาที่มีผู้ใช้งาน 2 รายโดยมีวัตถุประสงค์เพื่อให้ได้อัตราบิตต่ำสุดสูงที่สุดโดยนำเสนอการแปลงส่วนต่าง ๆ ของระบบโนมาไปเป็นองค์ประกอบของวิธี Q-Learning ได้แก่ เอเจนท์ แอคชัน สเตจ รางวัล และสภาพแวดล้อมซึ่งมีความสำคัญต่อกระบวนการเรียนรู้ ผลการจำลองระบบแสดงให้เห็นการจัดสรรกำลังด้วยวิธี Q-Learning มีการเรียนรู้เพื่อเพิ่มรางวัลในแต่ละเสตจ ในส่วนของสมรรถนะของระบบนั้นวิธี Q-Learning ให้อัตราบิตของผู้ใช้ทั้งสองรายใกล้เคียงกันและยังใช้ต่ำสุดที่สูงกว่าวิธีการจัดสรรกำลังที่มีอยู่ก่อนหน้าและเครื่องมือในไลบรารี่ของภาษา Python

คำสำคัญ: โนมา; การจัดสรรกำลัง; Q-Learning

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

บทความวิจัย

On Applying Q-Learning to Optimize Power Allocation in 2-users NOMA System

Phetnakorn Aermsa-Ard, Chonticha Wangsamad and Kritsada Mamat*

Department of Electronic Engineering Technology, College of Industrial Technology,
King Mongkut's University of Technology North Bangkok

* Corresponding author, E-mail: kritsada.m@cit.kmutnb.ac.th

Received: 1 August 2022; Revised: 28 November 2022; Accepted: 20 January 2023
Online Published: 13 April 2023

Abstract: This article considers a power domain non-orthogonal multiple access (NOMA) system which is a multiple access technique considered to be used in the 5G technology and beyond. Successive interference cancellation (SIC) is applied to decode user's signals and power allocation significantly affects the system performance. In this article, we propose to apply Q-learning which is one of the machine learning methods to solve a transmit power allocation problem in a 2-users NOMA system where the objective function is to maximize the minimum transmission rate. We show how to transform NOMA system into O-Learning components namely agent, action, stage, reward, and environment which are very important for the learning process. Numerical results show that the Q-learning offers higher reward in each step. For the system performance, the bit rates of two users in the system are very close to each other when the Q-learning is applied. Furthermore, the Q-learning offers a higher minimum rate than that performed by dynamic power allocation methods in the literature and optimizers in Python's library.

Keywords: NOMA; power allocation; Q-Learning

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

บทความวิจัย

1. บทนำ

เทคโนโลยีการสื่อสารไร้สาย (Wireless Communication) เป็นเทคโนโลยีที่มีความสำคัญในโลกยุคปัจจุบันเป็นอย่างมาก เนื่องจากความสามารถในการเข้าถึงพื้นที่ต่าง ๆ ได้มากกว่าการสื่อสารที่ต้องใช้สายส่งสัญญาณ (Wireline Communication) ในปัจจุบันเทคโนโลยีสื่อสารไร้สายมีการพัฒนาจนมาถึงในยุคที่ 5 หรือ 5G ซึ่งมีความต้องการเข้าใช้งานและอัตราส่งที่สูงขึ้นอย่างมาก เพื่อที่จะตอบโจทย์ความต้องการข้างต้นในปัจจุบันได้มีแนวคิดการใช้งานเทคนิคการเข้าถึงหลายส่วนแบบไม่ตั้งฉาก (Non- Orthogonal Multiple Access: NOMA) หรือโนมามาเป็นส่วนหนึ่งของการสื่อสารยุคใหม่โดยมีงานวิจัยจำนวนมากได้แสดงให้เห็นว่าวิธีโนมานั้นมีประสิภาพเชิงสเปกตรัม (Spectral Efficiency) มากกว่าวิธีการเข้าถึงหลายทางแบบตั้งฉาก (Orthogonal Multiple Access: OMA) หรือโอมา [1, 2]

ในการสื่อสารโนมานั้นผู้ใช้สามารถเข้าใช้ช่องสัญญาณที่เวลาและความถี่เดียวกันโดยผู้ใช้แต่ละรายจะมีกำลังหรือการเข้ารหัสที่แตกต่างกัน [3] สำหรับการตรวจจับข้อมูลนั้นในโดเมนของกำลังมีการใช้วิธี Successive Interference Cancellation หรือ (SIC) และการจัดสรรกำลังมีผลต่อสมรรถนะของระบบ [4, 5] งานวิจัยของ El-Sayed et. al., [4] ได้นำเสนอวิธีการจัดสรรกำลังสองวิธี โดยวิธีแรกมีวัตถุประสงค์เพื่อทำให้กำลังรับของผู้ใช้แต่ละรายมีค่าเท่ากันในขณะที่วิธีที่สองมีวัตถุประสงค์เพื่อรับประกันคุณภาพของการสื่อสารของผู้ใช้ในระบบ 1 ราย งานวิจัยของKaaffah และ Iskandar [5] นำเสนอการจัดสรรกำลังแบบพลวัต (Dynamic) และแสดงให้เห็นว่าการจัดสรรกำลังดังกล่าวให้สมรรถนะที่ดีกว่าการจัดสรรกำลังแบบตายตัวโดยมีผลรวมของอัตราบิตเป็นฟังก์ชันวัตถุประสงค์ (Objective function)

ในปัจจุบันเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) หรือ การเรียนรู้ของเครื่อง (Machine Learning : ML) ได้รับความสนใจและมีการประยุกต์ใช้งานที่หลากหลาย ตัวอย่างเช่น ในอุตสาหกรรมการเกษตร อุตสาหกรรมการแพทย์และอุตสาหกรรมการขนส่ง หลักการของ ML คือการเรียนรู้ผ่านกระบวนการลองผิดลองถูก (Trial-and-Error) ของเอเจนท์ (Agent) ต่อสภาพแวดล้อม (Environment) เพื่อสร้างกระบวนการตัดสินใจที่เหมาะสม [6] โดยในงานวิจัยที่ผ่านมามีการประยุกต์ใช้งาน ML ในระบบสื่อสารไร้สาย ตัวอย่างเช่น งานวิจัยของ Chen et. al., [7] ได้นำเสนอการใช้งาน ML เพื่อการหาค่าอัพลิงค์และดาวน์ลิงค์ (uplink and downlink) ที่เหมาะสมในระบบดีคัปลิง (Decoupling) และงานวิจัยของ Sun et. al., [8] ได้นำเสนอการนำ ML มาใช้เพื่อจัดสรรกำลังในระบบสื่อสารไร้สายโดยมีวัตถุประสงค์เพื่อทำให้อัตราบิตรวมมีค่าสูงที่สุด

บทความนี้นำเสนอการประยุกต์ใช้วิธี Q-Learning ซึ่งเป็นหนึ่งในวิธี ML เพื่อจัดสรรกำลังส่งในระบบโนมาที่มีผู้ใช้งานจำนวน 2 รายโดยมีวัตถุประสงค์เพื่อทำให้อัตราบิตต่ำสุดในระบบมีค่าสูงที่สุดโดยการกระทำดังกล่าวจะช่วยรับประกันคุณภาพสื่อสารของทั้งระบบ ในการประยุกต์ใช้ Q-Learning ในระบบโนมานั้นจำเป็นต้องมีการแปลงองค์ประกอบในระบบโนมาให้เป็นส่วนต่าง ๆ ใน Q-Learning ได้แก่ เอเจนท์ (Agent) สภาพแวดล้อม (Environment) เสตจ (State) แอคชันหรือการกระทำ (Action) และรางวัล (Reward) การ

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

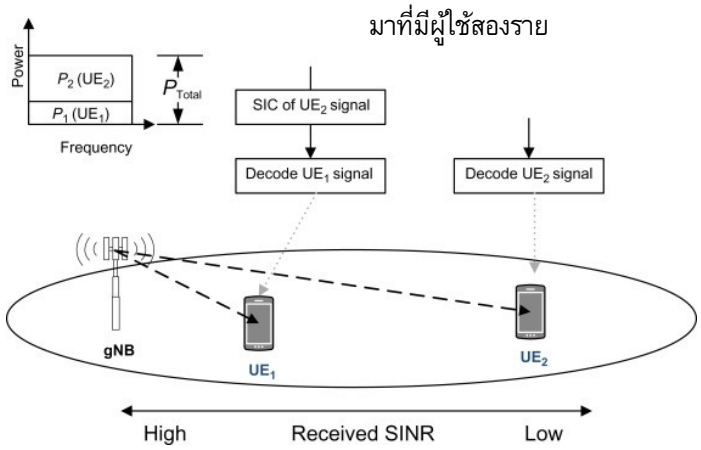
ประยุกต์ใช้ Q-Learning กับระบบที่ใช้ SIC นั้นเคยถูกพิจารณาในงานวิจัยของ Mete และ Girici [9] โดยในงานดังกล่าวได้ใช้ Q-Learning เพื่อการจัดสรรเวลาในการส่งข้อมูลเพื่อให้ได้จำนวนของแพ็คเกต (Packet) สูงที่สุด ผลการจำลองระบบแสดงให้เห็นว่าการประยุกต์ใช้ Q-Learning ให้อัตราบิตรวมที่ต่ำกว่าวิธีการจัดสรรกำลังที่มีอยู่ก่อนหน้าบางวิธีแต่ให้อัตราบิตต่ำสุดสูงที่สุด นอกจากนั้นแล้วบทความนี้ยังเปรียบเทียบวิธี Q-Learning กับเครื่องมือแก้ปัญหาค่าเหมาะสม (Optimizer) ที่มีอยู่ในไลบรารี่ของภาษา Python และพบว่าวิธี Q-Learning ให้อัตราบิตต่ำสุดสูงที่สุดเช่นกัน

บทความนี้มีส่วนประกอบตามเนื้อหาในแต่ละหัวข้อดังนี้ หัวข้อที่ 2 กล่าวถึงทฤษฎีพื้นฐานของโนมาและ Q-Learning รวมทั้งวิธีการจัดสรรกำลังที่มีอยู่ก่อน หัวข้อที่ 3 นำเสนอการประยุกต์ใช้ Q-Learning สำหรับแก้ปัญหาการจัดสรรกำลังในระบบโนมา หัวข้อที่ 4 นำเสนอผลการดำเนินการและการวิเคราะห์ผลที่ได้ หัวข้อที่ 5 เป็นการสรุปและอภิปรายผลการดำเนินการรวมทั้งนำเสนอแนวทางการพัฒนางานในอนาคต

2. ทฤษฎีที่เกี่ยวข้อง

2.1 ช่องสัญญาณโนมาและการจัดสรรกำลังส่ง

การเข้าถึงหลายส่วนแบบไม่ตั้งฉาก (Non-Orthogonal Multiple Access : NOMA) หรือโนมานั้นเป็นวิธีการเข้าใช้ช่องสัญญาณเมื่อมีผู้ใช้เป็นจำนวนมากในระบบสื่อสารไร้สายยุคที่ 5 และหลังจากนั้น โดยการสื่อสารในยุคที่ 1 ถึงยุคที่ 4 นั้นใช้การเข้าถึงหลายส่วนแบบแบ่งความถี่ (Frequency Division Multiple Access : FDMA) การเข้าถึงหลายส่วนแบบแบ่งเวลา (Time Division Multiple Access : TDMA) การเข้าถึงหลายส่วนแบบแบ่งรหัส (Code Division Multiple Access: CDMA) และการเข้าถึงหลายส่วนแบบตั้งฉากทางความถี่ (Orthogonal Frequency Multiple Access : FDMA) ตามลำดับ หลักการของโนมานั้นคือการให้ผู้ใช้แต่ละรายในระบบเข้าใช้งานช่องสัญญาณโดยใช้ความเวลาและถี่ร่วมกันซึ่งการแบ่งแยกผู้ใช้แต่ละรายนั้นอาจกระทำได้ในโดเมน ของกำลัง (Power domain) หรือโดเมนของคำรหัส (Code domain) รูปที่ 1 แสดงตัวอย่างการแบ่งผู้ใช้งานในโดเมนของกำลังในระบบโนมาที่มีผู้ใช้สองราย



รูปที่ 1 แสดงตัวอย่างการแบ่งผู้ใช้งานในโดเมนของกำลังในระบบโนมาที่มีผู้ใช้สองราย [10]

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

เมื่อพิจารณารูปที่ 1 พบว่าในระบบมีผู้ใช้จำนวนสองรายคือโดยรายที่ 1 คือ UE1 และรายที่ 2 คือ UE2 โดยผู้ใช้ทั้งสองรายใช้ความถี่ร่วมกันเพื่อติดต่อกับสถานีฐาน gNB และใช้กำลังที่แตกต่างกันโดยผู้ใช้รายที่ 1 ใช้กำลัง P_1 และผู้ใช้รายที่ 2 ใช้กำลัง P_2 ตามลำดับ จากรูปพบว่าอัตราส่วนของกำลังต่อการแทรกสอดและสัญญาณรบกวน (Signal-to-Interference plus Noise Ratio : SINR) ซึ่งเป็นตัวบ่งชี้คุณภาพของสัญญาณของผู้ใช้ 1 มีมากกว่าผู้ใช้รายที่ 2 ดังนั้นในการจัดสรรกำลังจึงต้องจัดสรรกำลังของผู้ใช้รายที่ 2 ให้มากกว่าผู้ใช้รายที่ 1 ตามที่แสดงในรูปวิธีการดังกล่าวเรียกว่า Super position coding และสัญญาณส่งที่ออกจากสถานีฐานสามารถเขียนอธิบายได้ดังนี้

$$x = \sqrt{\alpha_1 p_{tot}} s_1 + \sqrt{\alpha_2 p_{tot}} s_2 \quad (1)$$

เมื่อ x แทนสัญญาณส่ง α_1 และ α_2 แทนตัวประกอบการจัดสรรกำลัง p_{tot} แทนกำลังส่งรวม s_1 และ s_2 แทนข้อมูลของผู้ใช้รายที่ 1 และ 2 ตามลำดับ สัญญาณที่รับได้ของผู้ใช้สามารถเขียนอธิบายได้ดังนี้

$$y_k = c_k h_k x + n_k, k = 1,2 \quad (2)$$

เมื่อ y_k แทนสัญญาณรับของผู้ใช้ $c_k, 0 < c_k < 1$ แทนสัมประสิทธิ์การลดทอนของช่องสัญญาณ h_k แทนอัตราขยายของช่องสัญญาณและ n_k แทนสัญญาณรบกวนเกาส์สีขาวแบบบวก (Additive White Gaussian Noise : AWGN) ตามลำดับ สำหรับการตรวจจับข้อมูล (Decode) ของวิธีโนมานั้นใช้วิธี Successive Interference Cancellation หรือ (SIC) โดยวิธี SIC นั้นเริ่มจากการตรวจจับข้อมูลของผู้ใช้รายที่ 1 ก่อนโดยกำหนดให้ผู้ใช้รายที่ 2 ประพฤติตัวเป็นสัญญาณแทรกสอดของผู้ใช้รายที่ 1 เมื่อได้ข้อมูลของผู้ใช้รายที่ 1 แล้วจึงนำข้อมูลของผู้ใช้รายที่ 1 ลบออกจากสัญญาณรับของผู้ใช้รายที่ 2 ดังนั้นค่า SINR ของผู้ใช้รายที่ 1 สามารถเขียนอธิบายได้ดังนี้

$$\gamma_1 = \frac{\alpha_1 P_{tot} c_1^2 |h_1|^2}{\alpha_2 P_{tot} c_2^2 |h_2|^2 + \sigma_n^2} \quad (3)$$

และสำหรับผู้ใช้รายที่ 2

$$\gamma_2 = \frac{\alpha_2 P_{tot} c_2^2 |h_2|^2}{\sigma_n^2} \quad (4)$$

เมื่อ γ_1 และ γ_2 แทน SINR ของผู้ใช้รายที่ 1 และ 2 ตามลำดับ จาก SINR ของผู้ใช้ทั้งสองราย สามารถหาอัตราส่ง (Bitrate) ในหน่วยบิตต่อวินาทีต่อความถี่ (bit/sec/Hz) ของผู้ใช้แต่ละรายได้ดังนี้

$$R_k = log_2(1 + \gamma_k), k = 1,2 \quad (5)$$

และอัตราส่งรวม (Sum rate) ได้ดังนี้

$$R_{sum} = R_1 + R_2 \quad (6)$$

เมื่อพิจารณาสมการ (3)-(6) พบว่าสมรรถนะของระบบขึ้นอยู่กับตัวประกอบการจัดสรรกำลัง α_1 และ α_2 โดยงานวิจัย [4] ได้นำเสนอวิธี Channel inversion โดยวิธีดังกล่าวจะจัดสรรกำลังให้กับผู้ใช้แต่ละรายผกผันกับคุณภาพของช่องสัญญาณดังนี้

$$\alpha_k = \frac{1}{c_k^2 |h_k|^2 \sum_{i=1}^{K} \frac{1}{c_i^2 |h_i|^2}} \quad (7)$$

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

เมื่อพิจารณาสมการ (7) พบว่าวิธี Channel inversion จะจัดสรรกำลังที่มากกว่าให้กับผู้ใช้ที่มีคุณภาพช่องสัญญาณแย่กว่า นอกจากวิธี Channel inversion แล้วงานวิจัย [4] ยังได้นำเสนอวิธีจัดสรรกำลังที่รับประกันคุณภาพของการสื่อสาร (Quality of Service : QoS) ของผู้ใช้จำนวน 1 ราย ซึ่งโดยทั่วไปแล้วเป็นผู้ใช้ที่มีคุณภาพของช่องสัญญาณแย่การจัดสรรกำลังด้วยวิธีดังกล่าวสามารถเขียนอธิบายได้ดังนี้

$$\alpha_{itd} = \frac{\gamma(c_{itd}^2|h_{itd}|^2+\frac{\sigma_n^2}{P_{tot}})}{c_{itd}|h_{itd}|^2(1+\gamma)} \quad (8)$$

เมื่อ $\gamma < c_{itd}^2|h_{itd}|^2 \frac{\sigma_n^2}{P_{tot}}$ แทนค่า SINR ของผู้ใช้ที่ต้องการรับประกันคุณภาพของการสื่อสารและดรรชนี itd แทนผู้ใช้รายดังกล่าวสำหรับการจัดสรรกำลังวิธีนี้ ผู้ใช้รายอื่นในระบบจะได้รับการจัดสรรกำลังเท่ากันที่ค่า $(1-\alpha_{itd})/K$ เมื่อ K แทนจำนวนผู้ใช้ทั้งหมดในระบบ เมื่อพิจารณาวิธีการจัดสรรกำลังส่งทั้งสองวิธี พบว่าวิธีทั้งสองมีการปรับกำลังให้เหมาะสมกับคุณภาพของช่องสัญญาณของผู้ใช้งานเพื่อรักษาคุณภาพการของการสื่อสาร อย่างไรก็ตามทั้งสองวิธีไม่ได้คำนึงถึงคุณภาพการสื่อสารของผู้ใช้ที่มีอัตราบิตต่ำที่สุด

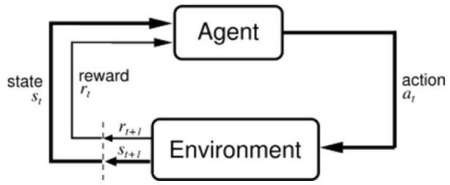
2.2 วิธี Q-Learning

วิธี Q-Learning จัดเป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning : ML) วิธีหนึ่งโดยวิธีการดังกล่าวเป็นการเรียนรู้จากการลองผิดลองถูก (trial-and-error) เพื่อหาวิธีการแก้ปัญหาที่เหมาะสมโดยหลักการทำงานของวิธี Q-Learning สามารถอธิบายได้ ดังรูปที่ 2 พบว่าวิธี Q-Learning มีส่วนประกอบที่สำคัญคือเอเจนท์ (Agent) สภาพแวดล้อม (Environment) เสตจ (State) การกระทำหรือแอคชัน (Action) และรางวัล (Reward) โดยการทำงานของวิธี Q-Learning นั้นอยู่บนพื้นฐานของกระบวนการตัดสินใจมาร์คอฟ (Markov Decision Process : MDP) เริ่มต้นจากเอเจนท์ดำเนินการแอคชัน a_t กับสภาพแวดล้อมที่เสตจปัจจุบัน s_t หลังจากนั้นสภาพแวดล้อมจะคืนค่ารางวัล r_t และย้ายไปที่เสตจ s_{t+1} จากนั้นเอเจนท์จะดำเนินการแอคชัน a_{t+1} เพื่อรับรางวัล r_{t+1} การกระทำดังกล่าวจะดำเนินการไปเรื่อย ๆ เพื่อหาแอคชันที่เหมาะสมสำหรับแต่ละเสตจ โดยความหมายของแอคชันที่เหมาะสมคือแอคชันที่กระทำต่อสภาพแวดล้อมในเสตจปัจจุบันแล้วได้รางวัลคืนค่ากลับมาสูงที่สุด เพื่อให้กระบวนการดังกล่าวเป็นไปอย่างถูกต้องจำเป็นต้องมีการสร้างตารางเพื่อบันทึกค่า Q (Q-value) สำหรับทุกเสตจและแอคชันทั้งหมดที่เป็นไปได้ตารางดังกล่าวคือ Q-table โดยจำนวนของเซลล์ใน Q-table ทั้งหมดเท่ากับจำนวนเสตจทั้งหมดคูณกับจำนวนแอคชันทั้งหมด สำหรับการอัพเดทค่า Q ใน Q-table นั้นใช้สมการของ Bellman ดังนี้ [6]

$$\begin{aligned} Q^{new}(s_t,a_t) \leftarrow & Q^{old}(s_t,a_t) \\ & + \eta((r(s_t,a_t) \\ & + \beta \max_a Q^{new}(s_{t+1},a) \\ & - Q^{old}(s_t,a_t)) \end{aligned} \quad (9)$$

โดยที่ η คืออัตราการเรียนรู้ (Learning rate) และ β แทนปัจจัยส่วนลด (Discount factor) เมื่อพิจารณาสมการ (9) พบว่าค่าที่อัพเดทใน Q-table หรือ $Q^{new}(s_t,a_t)$ ขึ้นอยู่กับค่าเดิม $Q^{old}(s_t,a_t)$ รางวัล $r(s_t,a_t)$ และรางวัลสูงสุดที่คาดหวัง $\max_a Q^{new}(s_{t+1},a)$ โดยเมื่อการอัพเดท Q-table เสร็จสมบูรณ์แล้วเอเจนท์ก็จะทราบแอคชันที่เหมาะสมสำหรับแต่ละเสตจ

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004



รูปที่ 2 หลักการทำงานของวิธี Q-Learning [11]

3. วิธีการดำเนินการ

บทความนี้นำเสนอการประยุกต์ใช้วิธี Q-Learning เพื่อแก้ปัญหาการจัดสรรกำลังในระบบโนมาเพื่อทำให้อัตราบิตของผู้ใช้ที่ต่ำสุดมีค่าสูงสุดเมื่อกำหนดให้ผู้ใช้ในระบบมีจำนวนสองราย โดยปัญหาดังกล่าวสามารถเขียนเป็นสมการได้ดังนี้

$$\max_{\{\alpha_1,\alpha_2\}} \gamma_{min}$$
$$\text{subject to } \gamma_k \geq \gamma_{min} \text{ for } k = 1,2$$
$$\alpha_1 + \alpha_2 \leq 1$$
$$\alpha_1, \alpha_2 > 0 \qquad (10)$$

เมื่อ γ_{min} แทนค่า SINR ต่ำที่สุด ปัญหาในสมการ (10) สามารถแก้ได้ด้วยเครื่องมือ (Optimizer) อื่นเช่น วิธี KKT (Karush–Kuhn–Tucker) ถ้าหากทราบค่าทางสถิติของช่องสัญญาณเช่นฟังชันความหนาแน่นของความน่าจะเป็น (probability density function : pdf) อย่างไรก็ตาม ถ้าหากไม่ทราบค่าทางสถิติของช่องสัญญาณปัญหาในสมการ (10) สามารถแก้ได้ด้วยการลองผิดลองถูกโดยในงานนี้เลือกใช้วิธี Q-Learning เนื่องจากมีกระบวนการทำงานที่เข้าใจง่ายและให้ผลเฉลยที่เหมาะสมซึ่งเป็นการเริ่มต้นที่ดีในการประยุกต์ใช้ ML เพื่อแก้ปัญหาการจัดสรรทรัพยากรในระบบสื่อสารไร้สาย

ในการแก้ปัญหาในสมการ (10) ด้วยวิธี Q-Learning นั้นเริ่มต้นได้กำหนดส่วนประกอบในระบบโนมาให้เป็นส่วนประกอบใน Q-Learning ดังนี้

- **เอเจนท์**: สถานีฐานซึ่งทำหน้าที่จัดสรรกำลังส่งให้ผู้ใช้แต่ละราย
- **แอคชัน**: การปรับค่า α_1 ให้เพิ่มขึ้นหรือลดลงครั้งละ 0.005 โดย α_2 สามารถคำนวณได้จาก $\alpha_2 = (1 - \alpha_1)$
- **เสตจ**: เมื่อสถานีฐานจัดสรรกำลังส่งให้แก่ผู้ใช้ทั้งสองรายจะสามารถคำนวณอัตราการส่งข้อมูลของผู้ใช้แต่ละรายได้ ดังนั้นเสตจคืออัตราการข้อมูลของผู้ใช้ทั้งสองราย R_1 และ R_2
- **รางวัล**: เนื่องจากต้องการทำให้อัตราการส่งข้อมูลต่ำสุดมีค่าสูงที่สุดซึ่งจะเป็นจริงได้ก็ต่อเมื่ออัตราการส่งข้อมูลของผู้ใช้ทั้งสองรายมีค่าใกล้เคียงกันมากที่สุดดังนั้นการกำหนดรางวัลให้แก่ระบบจึงใช้เปอร์เซ็นต์ความแตกต่างของอัตราการส่งข้อมูลของผู้ใช้ทั้งสองรายดังนี้

$$r_t = 100 - \frac{|R_1 - R_2|}{R_1} x100 \qquad (11)$$

เนื่องจากวิธี Q-Learning ต้องการรางวัลที่สูงดังนั้นจึงกำหนดให้รางวัลเท่ากับ 100 ลบ ด้วยเปอร์เซ็นต์ความแตกต่างตามที่แสดงในสมการ (11)

- **สภาพแวดล้อม**: เพื่อให้สอดคล้องกับเอเจนท์แอคชันและรางวัลก่อนหน้านี้ กำหนดสภาพแวดล้อมเอาไว้ตามที่แสดงใน **Algorithm 1**

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

Algorithm 1: สภาพแวดล้อม: $\boldsymbol{env}(\cdot)$

1. กำหนดเสตจปัจจุบัน s_t , α_1
2. เพิ่มหรือลด α_1 ครั้งละ 0.005 คำนวณ $\alpha_2 = (1-\alpha_1), R_1$ และ R_2
3. ไปยังเสตจถัดไป s_{t+1}
4. คำนวณรางวัลของเสตจปัจจุบัน $r(s_t, a_t)$ ด้วยสมการ (11)
10. **return** s_{t+1}และ $r(s_t, a_t)$

สำหรับการสร้าง Q-table ให้สอดคล้องกับการจัดสรรกำลังในระบบโนมาในสมการ (10) นั้นสามารถทำได้ตามที่แสดงไว้ในตารางที่ 1

ตารางที่ 1 Q-table ของการจัดสรรกำลังในระบบโนมาที่มีผู้ใช้ 2 ราย

เสตจ R_1	เสตจ R_2	เพิ่ม α_1 0.005	ลด α_1 0.005
ขั้นที่ 1	ขั้นที่ 1		
ขั้นที่ 2	ขั้นที่ 2		
.			
.			
.			
ขั้นที่ 10	ขั้นที่ 10		

เมื่อพิจารณาตารางที่ 1 พบว่าจำนวนขั้นของเสตจ R_1 และ R_2 มีค่าเท่ากับ 10 ทั้งนี้เนื่องจากอัตราบิตของผู้ใช้ทั้งสองรายจาก **Algorithm 1** เป็นค่าที่ไม่รู้จัก ดังนั้น จึงทำการแบ่งนับ (Quantize) ค่าที่ไม่รู้จักให้เป็นค่าที่รู้จักจำนวน 10 ขั้นมีค่าอยู่ในช่วง (0,8] โดยในตอนต้นนั้นได้กำหนดให้ทุกค่าใน Q-table เท่ากับ 0 ในส่วนของการอัพเดท Q-table นั้นได้ใช้วิธีการเรียนรู้ (Training) ตามที่แสดงใน **Algorithm 2**

Algorithm 2: กระบวนการเรียนรู้ (Training)

1. กำหนด จำนวนรอบ $N, \varepsilon, \eta, \beta$
2. **for** episode = 1:N **do**
3. **if** episode < $N/2$
4. $\phi = 0.5$
5. **else**
6. $\phi = \varepsilon$
7. **endif**
8. สุ่มค่า $0 < \alpha_1 < 1$
9. $s_{t+1}, r(s_t, a_t) = \boldsymbol{env}(\alpha_1)$
10. **while** $r(s_t, a_t) < 90$ **do**
11. if $rand() < \phi$ **then**
12. สุ่มทำแอคชัน
13. **else**
14. เลือกแอคชันจาก $argmax_a Q(s_t, a)$
15. **endif**
16. อัพเดทค่า Q ด้วยสมการ (9)
17. **endwhile**
18. **endfor**

โดยการทำงานใน **Algorithm 2** เริ่มจากการกำหนดรอบของการวนซ้ำสำหรับการเรียนรู้โดยในครึ่งแรกของการเรียนรู้นั้นเป็นการสุ่มทำแอคชันเพื่อให้มีค่า Q มาแทนค่า 0 ที่กำหนดไว้ตอนต้นใน Q-table ทุกค่า สำหรับครึ่งหลังนั้นเป็นการเลือกแอคชันที่ให้ค่า Q มากที่สุดในส่วนนี้แสดงไว้ในขั้นตอนที่ 3-7 สำหรับการวนซ้ำแต่ละรอบเริ่มต้นเป็นการสุ่มค่า α_1 และผ่านค่า α_1 เข้าสู่สภาพแวดล้อมตามขั้นตอนที่ 9 ในขั้นตอนที่ 10-17 เป็นการหาค่า α_1 ที่เหมาะสมเพื่อให้ได้รางวัลที่มีค่ามากกว่าหรือเท่ากับ 90 โดยระหว่างการ

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

ดำเนินการดังกล่าวค่าใน Q-table ก็จะมีการอัพเดทไปพร้อมกันด้วย ฟังก์ชัน $rand()$ ในบรรทัดที่ 11 เป็นการค่าสุ่มแบบยูนิฟอร์มที่มีค่าอยู่ในช่วง [0,1] ดังนั้นเมื่อค่าใน Q-table มีการอัพเดทอย่างเหมาะสมแล้ว สถานีฐานที่มีหน้าที่จัดสรรกำลังก็จะทราบการจัดสรรกำลังที่เหมาะสมสำหรับทุกเสตจ

4. ผลการดำเนินงาน

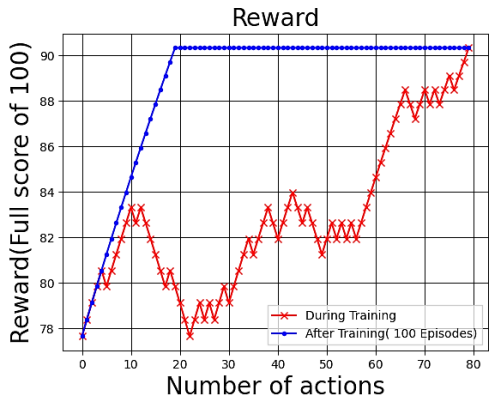
การทดสอบสมรรถนะของวิธี Q-Learning ในบทความนี้ใช้การเขียนโปรแกรมภาษา Python ในการจำลองระบบสื่อสารโนมาโดยการจำลองระบบนั้นใช้วิธีการสุ่มสร้างช่องสัญญาณเพื่อใช้คำนวณอัตราส่งและเป็นข้อมูลสำหรับการเรียนรู้ให้แก่วิธี Q-Learning โดยงานนี้ได้ทำการเปรียบเทียบข้อดีข้อเสียของวิธี Q-Learning กับวิธีการจัดสรรกำลังที่มีอยู่ก่อนหน้าและเครื่องมือในไลบรารี่ของภาษา Python

รูปที่ 3 แสดงรางวัลของวิธี Q-Learning ในระหว่างการเรียนรู้แสดงด้วยเส้นสีแดงและหลังจากเรียนรู้เรียบร้อยแล้วแสดงด้วยเส้นสีน้ำเงินโดยเมื่อพิจารณารางวัลในระหว่างการเรียนรู้นั้นจะพบว่าในแต่ละรอบของการเรียนรู้รางวัลที่ได้ในแต่ละรอบมีการเพิ่มขึ้นหรือลดลงทั้งนี้เนื่องจากในตอนเริ่มต้นค่าใน Q-table ยังมีค่าเป็น 0 หรือมีการอัพเดทเพียงไม่กี่ครั้ง อย่างไรก็ตามหลังจากเรียนรู้ไปได้ระยะหนึ่งค่าใน Q-table มีการอัพเดทเป็นค่าที่เหมาะสมดังจะเห็นได้จากรางวัลที่มีการเพิ่มเพียงอย่างเดียวเมื่อจำนวนแอคชันมีค่าตั้งแต่ 55 เป็นต้นไป ในส่วนของหลังการเรียนรู้นั้นพบว่ารางวัลมีการเพิ่มขึ้นเพียงอย่างเดียวและจะเพิ่มขึ้นจนถึงจุดสูงสุดเมื่อจำนวนแอคชันผ่านไปเพียง 19 เท่านั้น ทั้งนี้เนื่องจากหลังการเรียนรู้ค่าใน Q-table มีการอัพเดทอย่างเหมาะสมแล้ว

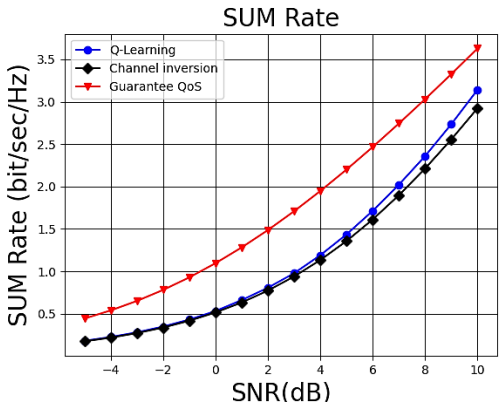
รูปที่ 4 แสดงการเปรียบเทียบวิธีสมรรถนะของการจัดสรรกำลังด้วยวิธี Q-Learning แสดงด้วยเส้นสีน้ำเงินกับวิธีที่มีอยู่ก่อนหน้าคือวิธี Channel inversion ที่มีการจัดสรรกำลังให้แก่ผู้ใช้ที่มีคุณภาพช่องสัญญาณแย่มากกว่าผู้ใช้ที่มีคุณภาพช่องสัญญาณดีกว่าแสดงด้วยเส้นสีดำและวิธีที่รับประกันคุณภาพการสื่อสารของผู้ใช้ 1 รายแสดงด้วยเส้นสีแดงโดยเปรียบเทียบ ในเทอมของอัตราบิตรวมของผู้ใช้ทั้งสองราย เมื่อพิจารณารูปดังกล่าวพบว่าวิธีที่รับประกันคุณภาพของการสื่อสารของผู้ใช้ 1 รายให้อัตราบิตรวมมากที่สุดในขณะที่วิธี Q-Learning ให้อัตราบิตรวมที่เทียบเท่ากับวิธี Channel inversion เมื่อค่า SINR มีค่าน้อย อย่างไรก็ตามเมื่อค่า SNR มีค่าเพิ่มขึ้นวิธี Q-Learning ให้สมรรถนะที่สูงกว่าวิธี Channel inversion และความแตกต่างของทั้งสองวิธีก็มีแนวโน้มเพิ่มมากขึ้นเมื่อค่า SNR มากขึ้น

รูปที่ 5 แสดงการเปรียบเทียบอัตราบิตต่ำสุดของระบบโนมาโดยเส้นสีน้ำเงินแทนวิธี Q-Learning เส้นสีดำแทนวิธี Channel inversion และเส้นสีแดงแทนวิธีที่รับประกันคุณภาพการสื่อสารของผู้ใช้ 1 ราย เมื่อพิจารณารูปดังกล่าวพบว่าวิธี Q-learning ให้อัตราบิตต่ำสุดสูงที่สุดในสามวิธีการจัดสรรกำลังเมื่อ SNR มีค่าสูงและค่าความแตกต่างดังกล่าวมีแนวโน้มเพิ่มขึ้นเมื่อ SNR มีค่ามากขึ้น โดยที่ค่า SNR มีค่าต่ำนั้นวิธี Q-Learning มีสมรรถนะเทียบเท่าวิธี Channel inversion

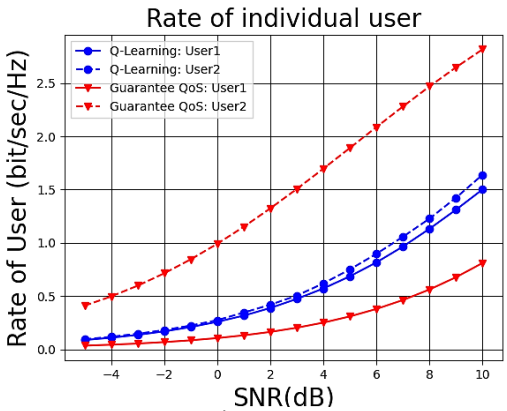
วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004



รูปที่ 3 แสดงรางวัลของวิธี Q-Learning ในระหว่างการเรียนรู้และหลังจากเรียนรู้เรียบร้อยแล้ว



รูปที่ 4 แสดงการเปรียบเทียบวิธีสมรรถนะของการจัดสรรกำลังด้วยวิธี Q-Learning กับวิธีที่มีอยู่ก่อนหน้า

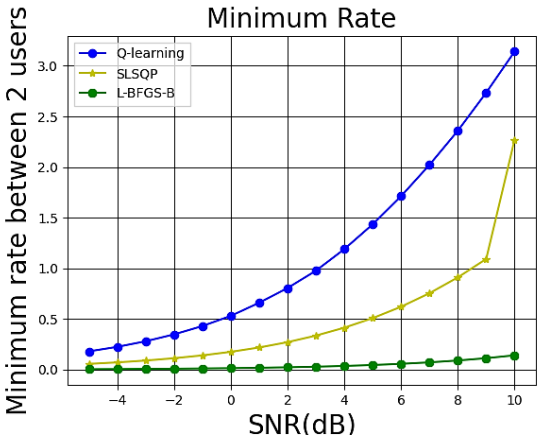


รูปที่ 5 แสดงการเปรียบเทียบอัตราบิตต่ำสุดของระบบโนมาของการจัดสรรกำลังทั้งสามวิธี

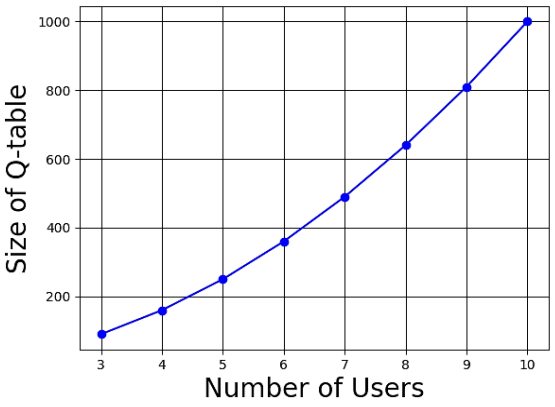
วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

เมื่อพิจารณารูปที่ 4 และ 5 พบว่าวิธี Q-Learning ให้อัตราบิตต่ำสุดที่สูงที่สุดและอัตราบิตรวมที่สูงกว่าวิธี Channel inversion อย่างไรก็ตามวิธี Q-Learning มีอัตราบิตรวมที่ต่ำกว่าวิธีที่รับประกันคุณภาพการสื่อสารของผู้ใช้ 1 ราย รูปที่ 6 แสดงอัตราบิตของผู้ใช้ทั้งสองรายในระบบโนมาจากการจัดสรรกำลังด้วยวิธี Q-Learning และวิธีที่รับประกันคุณภาพการสื่อสารของผู้ใช้ 1 ราย โดยเส้นสีน้ำเงินแทนวิธี Q-Learning และเส้นสีแดงแทนวิธีที่รับประกันคุณภาพการสื่อสารของผู้ใช้ 1 ราย เมื่อพิจารณารูปดังกล่าวพบว่าอัตราบิตของผู้ใช้รายที่ 1 และผู้ใช้รายที่ 2 จากวิธีที่รับประกันคุณภาพการสื่อสารของผู้ใช้ 1 รายมีความแตกต่างกันมากซึ่งไม่เป็นผลดีต่อระบบสื่อสารเนื่องจากการสื่อสารที่ดีนั้นคุณภาพการสื่อสารของผู้ใช้แต่ละรายไม่ควรแตกต่างกันมาก ในส่วนของวิธี Q-Learning นั้นอัตราการส่งข้อมูลของผู้ใช้รายที่ 1 และผู้ใช้รายที่ 2 มีความใกล้เคียงกันมาก

นอกจากการเปรียบเทียบวิธี Q-Learning กับวิธีการจัดสรรกำลังที่มีอยู่ก่อนหน้าทั้งสองวิธีดังที่แสดงในตอนต้นของหัวข้อแล้ว บทความนี้ยังได้เปรียบเทียบวิธี Q-Learning กับเครื่องมืออื่นที่มีอยู่ในไลบรารี่ของภาษา Python ได้แก่ SLSQP และ L-BFGS-B โดยกำหนดให้เครื่องมือทั้งสองแก้ปัญหาที่ทำให้เปอร์เซ็นต์ความต่างของอัตราบิตของผู้ใช้ทั้งสองรายมีค่าต่ำที่สุดตามที่แสดงในรูปที่ 7 โดยเส้นสีน้ำเงินแทนวิธี Q-Learning เส้นสีเหลืองแทนวิธี SLSQP และเส้นสีเขียวแทนวิธี L-BFGS-B จากรูปดังกล่าวจะพบว่าวิธี Q-Learning ให้อัตราบิตต่ำสุดสูงที่สุดทั้งนี้เนื่องจากเครื่องมือ SLSQP และ L-BFGS-B มีการจัดกำลังโดยไม่มีการเรียนรู้ข้อมูลเหมือนกับวิธี Q-Learning



รูปที่ 7 แสดงการเปรียบเทียบวิธี Q-Learning กับเครื่องมือในไลบรารี่ของภาษา Python



รูปที่ 8 แสดงขนาดของ Q-Table เมื่อพิจารณาที่จำนวนผู้ใช้ตั้งแต่ 3-10 ราย

รูปที่ 8 แสดงขนาดของ Q-Table เมื่อพิจารณาที่จำนวนผู้ใช้ตั้งแต่ 3-10 ราย จากรูปพบว่าขนาดของ Q-Table มีการเพิ่มขึ้นอย่างเอกซ์โพเนน*เชียล*กับจำนวนของผู้ใช้งานในระบบโนมาก่อให้เกิดความซับซ้อนของกระบวนการเรียนรู้ในแง่ของเวลาที่เพิ่มขึ้น

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

5. สรุปผลการดำเนินงานและข้อเสนอแนะ

บทความนี้นำเสนอการประยุกต์ใช้วิธี Q-Learning ซึ่งเป็นวิธีหนึ่งใน Machine Learning เพื่อแก้ปัญหาการจัดสรรกำลังในช่องสัญญาณโนมาที่มีผู้ใช้จำนวนสองรายโดยมีวัตถุประสงค์เพื่อทำให้อัตราบิตต่ำสุดมีค่าสูงที่สุด ผู้เขียนได้นำเสนอวิธีการแปลงองค์ส่วนต่าง ๆ ในระบบโนมาให้เป็นองค์ประกอบของวิธี Q-Learning ได้แก่ เอเจนท์ สภาพแวดล้อม แอคชัน เสตจและรางวัลตามลำดับ ผู้เขียนได้เปรียบเทียบวิธี Q-Learning กับวิธีการจัดสรรกำลังของระบบโนมาที่มีอยู่ก่อนหน้า ผลการจำลองระบบแสดงให้เห็นว่าวิธี Q-Learning ไม่ได้ให้อัตราบิตรวมที่มากที่สุดแต่ให้อัตราบิตต่ำสุดที่สูงที่สุด นอกจากนี้ผู้เขียนยังได้เปรียบเทียบวิธี Q-Learning กับเครื่องมือที่มีอยู่ในไลบรารี่ของภาษา Python และพบว่าวิธี Q-Learning ยังคงให้อัตราบิตต่ำสุดสูงที่สุด ทั้งนี้เนื่องจากวิธี Q-Learning นั้นมีกระบวนการเรียนรู้ข้อมูลเพื่อหาวิธีการจัดสรรกำลังที่เหมาะสมในแต่ละเสตจ ถึงแม้ว่าวิธี Q-Learning สามารถแก้ปัญหาการจัดสรรกำลังในช่องสัญญาณโนมาได้เป็นอย่างดี อย่างไรก็ตามเราไม่สามารถมองข้ามปัญหาเรื่องความซับซ้อนอันเนื่องการการเรียนรู้ข้อมูลได้ งานในอนาคตสนใจที่จะลดปัญหาจากความซับซ้อนดังกล่าวรวมถึงการปนะยุกต์ใช้งาน Q-Learning ในการแก้ปัญหาที่ซับซ้อนมากกว่านี้ด้วย

6. เอกสารอ้างอิง

[1] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, A survey of non-orthogonal multiple access for 5G, IEEE Communication Surveys Tutorials, 2018, 20(3), 2294–2323.

[2] M. Zeng, A. Yadav, O.A. Dobre, G.I. Tsiropoulos, and H.V. Poor, On the sum rate of MIMO-NOMA and MIMO-OMA systems, IEEE Wireless Communication Letter, 2017, 6(4), 534–537.

[3] B. Makki, K. Chitti, A. Behravan and M.-S. Alouini, A survey of NOMA: Current status and open research challenges, IEEE Open Journal of the Communications Society, 2020 1, 179-189.

[4] M.M. El-Sayed, A.S. Ibrahim and M.M. Khairy, Power allocation strategies for non-orthogonal multiple access, International Conference on Selected Topics in Mobile & Wireless Networking (MoWNeT-Egytp 2016), Proceeding, 2016, 1-6.

[5] M.A.M. Kaaffah and I. Iskandar, Power allocation effect on capacity of single carrier power domain non-orthogonal multiple access (NOMA), 7th International Conference on Wireless and Telematics (ICWT-Indonesia 2021), Proceeding, 2021, 1-5.

[6] R.S. Sutton and A.G. Barto, Reinforcement Learning: An Introduction, 2nd Ed., MIT Press, Cambridge, Massachusetts, London, 2017.

[7] M. Chen, W. Saad and C. Yin, Optimized uplink-downlink decoupling in LTE-U networks: An echo state approach, IEEE International Conference on Communications (ICC-Malaysia 2016), Proceeding, 2016, 1-6.

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (Print): 1686-9869, ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2023.04.004

[8] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu and N. D. Sidiropoulos, Learning to optimize: training deep Neural networks for interference management, IEEE Transactions on Signal Processing, 2018, 6(20), 5438-5453.

[9] E. Mete and T. Girici, Q-Learning based scheduling with successive interference cancellation, IEEE Access, 2020, 8, 172034-172042.

[10] https://www.sciencedirect.com/topics/engineering/non-orthogonal-multiple-access. (Accessed on 18 July 2022)

[11] https://medium.com/@nutorbitx/reinforcement-learning. (Accessed on 18 July 2022)