การรีบาลานซ์ในบิทคอยน์สปอตและป้องกันความเสี่ยงด้วยสัญญาซื้อขายล่วงหน้าผ่านการเรียนรู้เชิงลึกแบบเสริมกำลัง

ธีรพันธ์ จันทร์ปราโมทย์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2565

REBALANCING IN BITCOIN SPOT AND HEDGING WITH FUTURES BY DEEP REINFORCEMENT LEARNING

TEERAPAN JANPRAMOTE

A Thematic Paper Submitted in Partial Fulfillment of the
Requirements for the Degree of Master of Engineering
Department of Big Data Engineering
College of Innovative Technology and Engineering
Dhurakij Pundit University
Academic Year 2022

ใบรับรองสารนิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่

หัวข้อสารนิพนธ์ การรีบาลานซ์ในบิทคอยน์สปอตและป้องกันความเสี่ยงด้วยสัญญาซื้อขายล่วงหน้าผ่านการเรียนรู้เชิงลึกแบบเสริมกำลัง

เสนอโดย ธีรพันธ์ จันทร์ปราโมทย์

สาขาวิชา วิศวกรรมข้อมูลขนาดใหญ่

อาจารย์ที่ปรึกษาสารนิพนธ์ ดร.ธนภัทร ฆังคะจิตร

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

(ดร.สรรพฤทธิ์ มฤคทัต) ประธานกรรมการ

(ดร.ธนภัทร ฆังคะจิตร) กรรมการที่ปรึกษาสารนิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น) กรรมการ

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ รับรองแล้ว

(ดร.ชัยพร เขมะภาตะพันธ์) คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

วันที่ 31 เดือน กรกฎาคม พ.ศ. 2566

หัวข้อสารนิพนธ์ การรีบาลานซ์ในบิทคอยน์สปอตและป้องกันความเสี่ยงด้วยสัญญาซื้อขายล่วงหน้าผ่านการเรียนรู้เชิงลึกแบบเสริมกำลัง
ชื่อผู้เขียน ธีรพันธ์ จันทร์ปราโมทย์
อาจารย์ที่ปรึกษา ดร.ธนภัทร ฆังคะจิตร
หลักสูตร วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา 2565

บทคัดย่อ

ในช่วงหลายปีที่ผ่านมาสกุลเงินดิจิทัล (Cryptocurrency) มีการพัฒนาไปสู่สินทรัพย์ทางการเงินที่มีการซื้อขายทั่วโลกเนื่องจากมีลงทุนที่ได้รับผลตอบแทนจำนวนมหาศาล เพื่อการลงทุนในสกุลเงินดิจิตอลอย่างมีประสิทธิผล จึงมีงานวิจัยจำนวนมากที่เสนอแนวทางในการลงทุนในสกุลเงินดิจิตัลโดยใช้เทคนิคทางด้านการเรียนรู้เชิงลึกแบบเสริมกำลัง (Reinforcement learning) เพื่อทำการสอนตัวแทน (Agent) ให้เรียนรู้พฤติกรรมของตลาดในหลากหลายรูปแบบ รวมถึงการปรับสมดุลย์พอร์ตการลงทุน (Rebalancing) เพื่อตอบโจทย์เป้าหมายการลงทุนและรักษาระดับความเสี่ยงที่เรายอมรับได้ แต่อย่างไรก็ตามมีเหตุการณ์ระบบของเหรียญสกุลเงินดิจิทัลถูกโจมตีอย่างต่อเนื่องส่งผลต่อความเชื่อมั่นและความผันผวนของค่าสกุลเงินดิจิทัล

ดังนั้นงานวิจัยนี้จึงนำเสนอการแนวทางการลงทุนในสกุลเงินดิจิทัลที่มีประสิทธิภาพโดยการป้องกันความเสี่ยง (Hedging) ด้วยการชอร์ตสัญญาซื้อขายล่วงหน้า (FUTURES) เพื่อลดการขาดทุนในกรณีที่สกุลเงินดิจิตัลอ่อนตัว งานวิจัยนี้ทำการปรับสมดุลย์พอร์ตการลงทุนระหว่าง เงินสด (USDT/USD), สกุลเงินดิจิตัล Bitcoin (BTC/USD) และ การขายชอร์ตสัญญาซื้อขายล่วงหน้า (Short Futures Contracts) ทั้งนี้ได้ใช้เทคนิคการเรียนรู้แบบเสริมกำลังแบบ Double Deep Q Network ร่วมกับ Muti-scale Continuous Loss เพื่อไม่ให้ตัวแทนกระทำการตัดสินใจผิดซ้ำ ซึ่งลดโอกาสการขาดทุนอย่างต่อเนื่องของพอร์ตการลงทุน

จากผลการทดลองในช่วงเดือน มิถุนายน 2022 ถึง เดือน มิถุนายน 2023 โดยกำหนดสัดส่วนของพอร์ตการลงทุนเป็นเงินสด 50%, Bitcoin 50% และป้องกันความเสี่ยง 95% ของบิทคอยน์ที่มีอยู่พบว่าให้ผลกำไรตอบแทนใกล้เคียงกับกลยุทธ์ซื้อ Bitcoin และถือไว้ (Buy & Hold) ทั้งนี้เนื่องจากเป็นช่วงตลาดขาขึ้น

อย่างไรก็ตามเมื่อพิจารณาถึงมาตรวัดอื่นที่เกี่ยวข้องกับความเสี่ยง Max Drawdown ลดลงมาก เมื่อเทียบกับการซื้อ Bitcoin และถือไว้ (Buy & Hold) นั่นคือ 10.53% กับ 35.43% ตามลำดับ และ ค่า Annualized Volatility เท่ากับ 20.35% ซึ่งมีค่าน้อยกว่าเมื่อเทียบกับความผันผวนของ Bitcoin ที่มีค่า 51.57% ในอนาคตอันใกล้ถ้าสามารถปรับเปลี่ยนสัดส่วนของการปรับสมดุลย์ในพอร์ตการลงทุนตามพฤติกรรมของตลาดสกุลเงินดิจิทัล ย่อมส่งผลต่อผลกำไรที่เพิ่มขึ้นของพอร์ตการลงทุน

คำสำคัญ : การลงทุนในสกุลเงินดิจิทัล, การปรับสมดุลย์พอร์ตการลงทุน, การลดความเสี่ยงในการลงทุน, การชอร์ตสัญญาซื้อขายล่วงหน้า, การเรียนรู้เชิงลึกแบบเสริมกำลัง

(อาจารย์ที่ปรึกษา)

A Thematic Paper Title	REBALANCING IN BITCOIN SPOT AND HEDGING WITH FUTURES BY DEEP REINFORCEMENT LEARNING
Author	TEERAPAN JANPRAMOTE
A Thematic Paper Advisor	Dr. Thanapat Kangkachit
Program	Big Data Engineering
Academic Year	2022

ABSTRACT

Recently, cryptocurrencies have evolved into a globally traded financial asset with huge amounts of returns. To effectively invest in cryptocurrencies, many studies applied deep reinforcement learning techniques to teach agents to learn a variety of market behaviors. In addition, rebalancing investment portfolios was utilized to meet investment goals and maintain the level of accepted risk. However, there are several attacks on the digital currency coin systems, affecting the confidence and volatility of the digital currency value.

In this research, we present an effective way to invest in cryptocurrencies by hedging on short Futures contracts to reduce losses in case of the decreased cryptocurrencies' values. Therefore, our investment portfolio comprises three types of assets; cash in USDT/USD, Bitcoin (BTC/USD) and short Futures contracts. We applied the double deep Q-networks with the multi-scale continuous loss (MSCL) to avoid repeating wrong decisions made by agents resulting in the reduced possibility of sustained losses.

The experimental data are collected during June 2022 and June 2023 which is considerably in the bull market. Initially, the portfolio contains 50% cash, 50% Bitcoin and 95% hedging of existing Bitcoin.

Although, our proposed method offers equivalent cumulative returns with the Buy and Hold (B&H) strategy. Contrastingly, our method outperforms B&H in the max drawdown and annualized volatility measures risk-related metrics. Considering the max drawdown and annualized volatility measures, our proposed portfolio offered only 10.53% and 20.35% compared to 35.43% and 51.75% by B&H. Furthermore, the rebalancing proportion of assets should be automatically adjusted according to cryptocurrency market behaviors to increase portfolio returns.

Keywords: Investing in Cryptocurrencies , Portfolio Rebalancing, Short Futures Contracts, Risk Reduction in Investing,Deep Reinforcement Learning

(Advisor)

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงได้ เพราะ ได้รับคำแนะนำ และการผลักดัน โดยอาจารย์ที่ปรึกษาสารนิพนธ์ ดร. ธนภัทร ฆังคะจิตร รวมถึงการสละเวลาให้คำแนะนำ ปรึกษา และ ไอเดีย ของการประยุกต์ใช้อัลกอริทึม ผู้เขียนจึงขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ขอขอบพระคุณ ผู้ช่วยศาสตรจารย์ ดร. ดวงใจ จิตคงชื่น ที่ได้เป็นกรรมการในการสอบสารนิพนธ์ และ กรุณาให้คำชี้แนะจนกระทั่งสารนิพนธ์ฉบับนี้เสร็จสมบูรณ์

สุดท้ายนี้ผู้เขียนขอบคุณผู้มีส่วนเกี่ยวข้องทุกท่านที่ทำให้สารนิพนธ์ชิ้นนี้สำเร็จ และสามารถนำไปพัฒนาต่อเพื่อใช้งานจริงได้ โดยหวังว่าจะเป็นประโยชน์ต่อผู้ลงทุนทุกท่าน

ธีรพันธ์ จันทร์ปราโมทย์

สารบัญ

สารบัญ (ต่อ)

บทที่ **หน้า**

สารบัญตาราง

ตารางที่ **หน้า**

สารบัญภาพ

ภาพที่ **หน้า**

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การลงทุนแบบ Automate เริ่มเข้ามามีบทบาทในตลาดเงินมากขึ้น ทั้งในกองทุน Hedge Fund ต่างๆรวมถึงนักลงทุนรายย่อยทั่วไป ซึ่งมีการให้บริการอยู่ในบาง Cryptocurrency Exchange การปรับสมดุลย์(Rebalance) ก็เป็นหนึ่งในบริการที่มีให้แต่ด้วยเหตุการณ์ที่เพิ่งเกิดขึ้นในตลาด Cryptocurrency เมื่อไม่นานมานี้ เช่น การถูกโจมตีของ Terra ที่สร้างเหรียญ LUNA และ Stable coin USDC หรือ การล่มสลายของ FTX เว็บ Cryptocurrency Exchange ที่มีผู้ใช้เป็นจำนวนมากซึ่งสะท้อนให้เห็นถึงความเสี่ยงต่างๆของการลงทุนใน Cryptocurrency

ซึ่งงานวิจัยของ Berend Jelmer Dirk Gort, Xiao-Yang Liu, Xinghang Sun, Jiechao Gao, Shuaiyu Chen, Christina Dan Wang (2022) [9] ใช้ ดัชนี CVIX (Cryptocurrency Volatility Index) เข้ามาเป็นตัวควบคุมความเสี่ยงโดยหาก CVIX สูงเกินไปจากระดับที่กำหนดไว้ก้จะหยุดการซื้อ แล้วขาย Cryptocurrency ที่มีอยู่ทิ้งทั้งหมดแล้วจะกลับมาทำการซื้อขายต่อเมื่อ CVIX กลับมาอยู่ในระดับปกติ Giorgio Lucarelli, Matteo Borrotti (2020) [10] มีการใช้การตัดขาดทุนเมื่อขาดทุนมากกว่า 2.3% และ ทำกำไรเมื่อกำไรมากกว่า 6.5% ในการซื้อขาย Cryptocurrency Gang Huang, Xiaohua Zhou and Qingyang Song (2020) [11] ประยุกต์ใช้ Deep Reinforcement Learning ในการซื้อขายหุ้นจีนในดัชนี CSI300 โดยให้ Agent สามารถเลือกขายชอร์ตหุ้น (Short Sell) เพื่อทำกำไรขาลงหากหุ้นถูกประเมิณว่ามีมูลค่าสูงเกินมูลค่าที่แท้จริง ซึ่ง Agent สามารถเลือกกระทำได้เพียงอย่างใดอย่างหนึ่งกับหุ้นว่าจะซื้อ หรือขายชอร์ต แต่เนื่องจากตลาด Cryptocurrency เป็นตลาดที่มีความผันผวนมาก การใช้ ดัชนี หรือตั้งจุดตัดขาดทุนอาจเกิดการขายตัดขาดทุนโดยไม่จำเป็นหากเป็นระดับที่ไม่เหมาะสม หรือประยุกต์ใช้ขายชอร์ตกับตลาด Cryptocurrency เพื่อทำกำไรในช่วงที่ราคาลง ก็มีโอกาสขาดทุนในพอร์ตการลงทุนหากมีสัดส่วนมากเกินไปในช่วงที่ราคาขึ้น

ดังนั้น ในการทำการปรับสมดุลย์ ผู้เขียนจึงเพิ่มการขายชอร์ตสัญญาซื้อขายล่วงหน้า แบบไม่มีวันหมดอายุเข้าไปในพอร์ตการลงทุน ทีทำการปรับสมดุลย์เหรียญ Bitcoin เพื่อป้องกันความเสี่ยงที่จะเกิดขึ้นกับพอร์ตการลงทุน โดยอัลกอริทึ่มที่นำมาใช้คือ Double Deep Q Network ที่ใช้การคิด Reward แบบ Multi-scale continuous loss (MSCL)เพื่อลดโอกาสการลดลงอย่างต่อเนื่องของมูลค่าสินทรัพย์สุทธิซึ่งมีแนวคิดมาจากการเพิ่มบทลงโทษที่มากขึ้นในแต่ละครั้งถ้าทำผิดแบบเดิมซ้ำโดยให้เริ่มต้นพอร์ตการลงทุนที่สัดส่วน Bitcoin spot 50%:เงินสด50%ผลลัพธ์ที่ได้คือ Max Drawdown ลดลงมากเมื่อเทียบกับการซื้อ Bitcoin แล้วถือ (Buy & Hold) นั่นคือ 10.53% กับ 35.43% ตามลำดับ และ Cumulative Return ที่ได้ก็แตกต่างกันน้อยมาก คือ 21.00% กับ 21.14% และ ทำการเปรียบเทียบกับอัตราส่วนเริ่มต้นแบบ Bitcoin spot 30%:

เงินสด 70% กับ Bitcoin spot 70%:เงินสด 30% เมื่อลดการถือ BTC ลงจะทำให้ Annualized Sharpe Ratio และ Calmar Ratio มีสัดส่วนสูงขึ้น การเพิ่ม BTC เพิ่มขึ้นจะทำให้ได้ Max Drawdown และ Volatility ลดลง เพราะมีการ ป้องกันความเสี่ยงเริ่มต้นที่ 95% ของ BTC ที่มี

1.2 วัตถุประสงค์ของงาน

1. เพื่อนำเสนอการจัดพอร์ตการลงทุนในการลงทุน Bitcoin ที่เพิ่มการขายชอร์ตสัญญาซื้อขายล่วงหน้าเข้ามาเพื่อป้องกันความเสี่ยง
2. เพื่อนำเสนอการปรับสมดุลย์ ด้วย Double Deep Q Network ที่ใช้การคิด Reward แบบ Multi-scale continuous- loss (MSCL) เพื่อป้องกันไม่ให้ Agent ทำผิดซ้ำแบบเดิม

1.3 ขอบเขตของงาน

1. พัฒนา trading bot ของ Bitcoin
2. การขายชอร์ตสัญญาซื้อขายล่วงหน้า และ Spot ที่ใช้จะเป็น BTC/USDT
3. การปรับสมดุลย์ เป็นแบบ daily
4. ใช้ Multi Asset Mode ซึ่งทาง Binance เปิดให้สามารถนำเอาทรัพย์สินใน พอร์ตการลงทุน ไปวางเป็นหลักประกันเพื่อเปิดการขายชอร์ตสัญญาซื้อขายล่วงหน้าได้
5. สัญญาซื้อขายล่วงหน้าเป็นแบบไม่มีวันหมดอายุ (Perpetual)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อศึกษาการประยุกต์ใช้เทคนิค Reinforcement learning กับ ด้านการลงทุน
2. เพื่อให้เป็นตัวอย่างหนึ่งของการปรับสมดุลย์ ในตลาด Cryptocurrency ที่เพิ่ม การขายชอร์ตสัญญาซื้อขายล่วงหน้า ใน พอร์ตการลงทุน ซึ่งสามารถช่วยลดการลดลงของ มูลค่าสินทรัพย์สุทธิ ได้

1.5 นิยามศัพท์

1. Double Deep Q Network หมายถึง เทคนิคหนึ่งในการเรียนรู้แบบเสริมแรง (Reinforcement learning) ที่ใช้ Network 2 อัน โดยอันนึงจะเลือก Action ส่วนอีกอันจะใช้เป็นการ Evaluate ของ Action ที่ถูกเลือกมา

2. Cryptocurrency คือ สินทรัพย์ดิจิทัล ประเภทหนึ่งที่มีการรักษาความปลอดภัยด้วยการเข้ารหัส ถูกออกแบบมาเพื่อใช้เป็นสื่อกลางในการแลกเปลี่ยนเช่นเดียวกับสกุลเงินทั่วไป (Fiat Currency) เพียงแต่ไม่สามารถจับต้องได้

บทที่ 2
แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

สารนิพนธ์ชิ้นนี้มีวัตถุประสงค์เพื่อนำ Reinforcement learning ด้วยวิธีการ Double Deep Q-Network มาใช้ในการปรับสมดุลย์ Cryptocurrency เช่น Bitcoin และ ตราสารอนุพันธ์แบบ Perpetual ใน Multi-Asset Mode เพื่อทำการ ป้องกันความเสี่ยง ลดความเสี่ยงที่จะเกิดขึ้นกับ พอร์ตการลงทุน จึงทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังนี้

2.1 Reinforcement learning

2.2 Double Q-learning

2.3 Deep Q Network

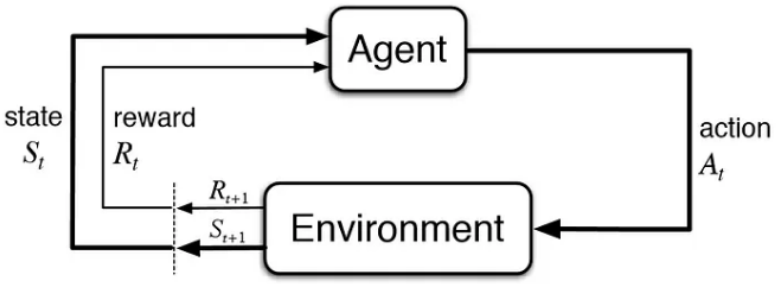
2.4 Double Deep Q Network

2.5 Technical Indicator

2.6 งานวิจัยที่เกี่ยวข้อง

2.1 Reinforcement learning

คือ วิธีการเรียนรู้ของเครื่องจักรวิธีหนึ่งซึ่งมีพื้นฐานอยู่บนการให้รางวัลบนพฤติกรรมที่น่าพอใจ และ/หรือ ลงโทษในพฤติกรรมที่ไม่พอใจ โดยทั่วไปแล้ว ผู้กระทำ สามารถรับรู้และตีความสภาพแวดล้อม แล้วลงมือกระทำ แล้วเรียนรู้ผ่านการลองผิดลองถูกเพื่อหา optimal policy ซึ่งก็คือควรกระทำสิ่งใดใน State นั้น เพื่อให้บรรลุเป้าหมายและได้รับรางวัลสูงที่สุด



ภาพที่ 2.1 รูปแบบการทำงานของ Reinforcement Learning [8]

องค์ประกอบของ Reinforcement learning จะประกอบไปด้วย

1. Agent คือ ผู้ตัดสินใจเลือกกระทำ (Action) ภายใต้สถาณการณ์ปัจจุบัน (State) เพื่อให้บรรลุเป้าหมายและได้รับรางวัลมากที่สุด

2. State คือ สถานะ หรือ สภาวะของระบบ ซึ่งเป็นผลจากการกระทำของ Agent โดย State จะถูกสร้างระบบ หรือ สภาพแวดล้อมภายนอก (Environment)

3. Environment คือ ระบบ หรือ สภาวะแวดล้อมภายนอกที่คอยรับการกระทำของ Agent หลังจากนั้นก็จะส่ง รางวัล (Reward) และ สถานะ State ถัดไปกลับไปให้กับ Agent

4. Reward คือ รางวัลที่ได้จากการกระทำ ใน แต่ละ State ซึ่ง เพื่อให้บรรลุเป้าหมายและได้รับรางวัลสูงที่สุด

5. Policy คือ วิธีการตัดสินใจของ Agent ในการเลือก Action

6. Value Function คือมูลค่าของ action หรือ state กล่าวคือถ้า action ใน state ใดมีมูลค่าสูง แสดงว่าหากเราทำ action ใน state นั้น แล้วเราจะมีโอกาสที่จะได้ reward ที่สูงตามมา

2.2 Double Q-learning

2.2.1 นิยามพอสังเขป

Double Q-learning ถูกเสนอเพื่อใช้แก้ปัญหาการการประมาณ Action value (Q-value) ที่มากเกินไปใน Q-learning ซึ่ง Agent เลือก Non-optimal Action ในสเตทนั้นๆ เพราะมีค่า Q-value สูงสุดใน Q-learning Optimal policy ของ Agent จะถูกเลือกจาก Action ที่ดีที่สุดใน State นั้น ซึ่งนั่นก็คือ มีค่า Q-value สูงที่สุด แต่อย่างไรก็ตาม Agent ไม่รู้จัก Environment ตั้งแต่แรกทำให้ต้องประมาณการ Q-value ในตอนแรกและค่อยๆปรับค่าในแต่ละรอบ วิธีนี้ทำให้มีค่า noise ที่สูงซึ่ง Action ที่ดีที่สุดมักมีค่าน้อยกว่าเมื่อเทียบกับ non-optimal action ในหลายๆกรณี ซึ่งก่อให้เกิด Overestimation ในการปรับค่า Q-value

สมการการปรับค่า Q-value ของ Q-learning หรือ Bellman Equation คือ

$$Q(s,a) = Q(s,a) + \alpha(r + \gamma\, max(Q(s',a')) - Q(s,a)) \quad (1)$$

เมื่อ Q(s,a) คือ Q function ของ state s และ action a

r คือ Reward ของ state s และ action a

γ คือ discount rate

α คือ learning rate ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1

Q(s′,a′) คือ Q function ของ state s′ และ action a′ ซึ่ง เป็น State ถัดไป

ซึ่ง Double Q-learning จะใช้ 2 action-value function เป็นตัวประมาณค่า ซึ่งจะถูกสร้างมาจาก Experience ที่ต่างกัน เช่น การอัพเดทค่าใน Q1 การเลือก optimal action ของ next state จะใช้(argmax Q1(s′,a)) แต่เราจะค่า Q-value ของ State และ Action นั้น จะใช้จากตาราง Q2 แทน สมการการ ของ Double Q-learning คือ

$$Q1(s,a) = Q1(s,a) + \alpha(r + \gamma Q2(s, \text{argmax}(Q1(s',a'))) - Q1(s,a)) \quad (2)$$

เมื่อ Q1(s,a) คือ Q function ตัวที่ 1 ของ state s และ action a

Q2(s,a) คือ Q function ตัวที่ 2 ของ state s และ action a

r คือ Reward ของ state s และ action a

γ คือ discount rate

α คือ learning rate ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1

Q1(s′,a′) คือ Q function ตัวที่ 1 ของ state s′ และ action a′ ซึ่ง เป็น State ถัดไป

ด้วยวิธีนี้ก็จะลดการปัญหาดังกล่าวลงได้

2.3 Deep Q Network

2.3.1 นิยามพอสังเขป

ถ้าState มีจำนวนมากการบันทึกค่าลงในตาราง Q-tables จะทำได้ช้าและยาก ดังนั้นจึงมีการใช้ Neural Network เข้ามาแทนในการประมาณค่า Q(s,a; θ) โดย θ คือค่า weights ใน Neural Network Deep Q Network จะใช้ State เป็น Input และ Output เป็น Action โดยในแต่ละรอบที่ทำการ Train จะมีการปรับค่า weights θ ในทุกๆรอบเพื่อทำให้ค่า Loss ซึ่งโดยทั่วไปจะใช้แบบ Mean Square Error ของระหว่าง ค่า Q-value จาก Bellman equation กับ Q-value ที่ประมาณจาก Neural Network มีค่าน้อยที่สุด สมการ Q-value จาก Bellman equation ที่ใช้เป็น Target คือ

$$y = r + \gamma \max Q(s', a'; \theta \mid s, a) \quad (3)$$

เมื่อ y คือ ค่าที่ใช้เป็น Target ใน Loss Function

r คือ Reward ของ state s และ action a

γ คือ discount rate

Q(s′,a′; θ) คือ ค่า Q-value ของ state s′ และ action a′ ซึ่ง เป็น State ถัดไป

และ Loss Function คือ

$$L(\theta) = (y - Q(s, a; \theta))^2 \quad (4)$$

เมื่อ y คือ ค่าที่ใช้เป็น Target จาก Bellman equation
Q(s,a;θ) คือ ค่า Q-value state s และ action a ที่ประมาณค่าจาก weight ใน Neural Network

โดยปกติแล้ว Neural Network จะมีปัญหาเรื่อง Correlation ในลำดับของ Training Data และโครงสร้างของ Deep Q Network ที่ใช้ค่า Weight θ เดียวกันในการหาค่า Q-Value ทั้ง Target และ ค่าที่พยากรณ์ แล้วก็จะ Update ค่า Weight θ ทำให้ Target เกิดการเปลี่ยนแปลงที่ไวเกินไปซึ่งส่งผลกับการ Train ได้ เพื่อแก้ปัญหานี้จึงต้องเพิ่มหัวข้อดังต่อไปนี้

(1) Experience Replay ข้อมูลประสบการณ์ของ Agent จะถูกเก็บไว้ในหน่วยความจำซึ่งประกอบด้วย (s,a,r,s') แล้วก็จะถูกสุ่มหยิบขึ้นมาเป็น Mini-Batches เพื่อ Train ตัว Neural Network ซึ่งช่วยลดปัญหา Correlation ของข้อมูลได้

(2) Target Network แทนที่จะใช้ Network เดียวในการพยากรณ์ค่า Q(s,a; θ) แต่จะเพิ่มตัว Target Network Q(s,a; θ') ซึ่งจะคัดลอกค่า Weight จาก θ จาก Online Network ในทุกๆ τ-Steps จากนั้นก็จะใช้ Target Network หาค่า Target ในสมการที่ (3)

$$y = r + \gamma \max Q(s', a'; \theta' \mid s, a) \quad (5)$$

ซึ่งสามารถลดปัญหาของค่า Target ที่เปลี่ยนแปลงไวเกินไปลงได้

2.4 Double Deep Q Network

2.4.1 นิยามพอสังเขป

Deep Q Network คำนวณค่า Target ด้วยการใช้ค่า Max ซึ่งคือการเลือกค่าที่มากที่สุด แล้วการประมาณค่า Q-value และ เลือก Action ของ Target ก็จะทำพร้อมกันบน Network เดียวกัน ซึ่งทำให้เกิดปัญา Overestimation ดังนั้น Double Deep Q Network จึงแยกขั้นตอนของการประมาณค่า Q-value และ เลือก Action ออกจากกัน โดยให้ การเลือก Action ของ Q(s',a') กระทำบน Online Network θ ส่วน การประมาณค่า Q-value กระทำบน Target Network θ' ซึ่งเขียนได้ดังสมการต่อไปนี้

$$y = r + \gamma\ Q(s', \text{argmax}\, Q(s', a'; \theta); \theta' \mid s, a) \quad (6)$$

2.5 Technical Indicators

2.5.1 Moving Average Convergence-Divergence (MACD) [5]

คือ Technical indicator ที่สร้างโดย Gerald Appeal ซึ่ง MACD สร้างมาจากส่วนต่างของ Fast EMA กับ Slow EMA

$$\text{MACD} = \text{EMA}_{12} - \text{EMA}_{26}$$

และ Signal Line ที่ใช้ค่าเฉลี่ย EMA 9 วันของ MACD

$$\text{Signal Line} = \text{EMA}_{9}$$

2.5.2 Even Better Sine Wave (EBSW) [6],[7]

คือ การดัดแปลงของ Hilbert sine wave ซึ่งใช้ดู Cycle ของราคาโดย EBSW มีค่าระหว่าง -1 ถึง 1 หากมีค่ามากกว่า 0.85 จะอยู่ในภาวะ Overbought หาก น้อยกว่า -0.85 จะอยู่ในภาวะ Oversold

```
{
  Even Better Sinewave Indicator
  © 2013  John F. Ehlers
}

Inputs:
  Duration(40);

Vars:
  alpha1(0),
  HP(0),
  a1(0),
  b1(0),
  c1(0),
  c2(0),
  c3(0),
  Filt(0),
  count(0),
  Wave(0),
  Pwr(0);

//HighPass filter cyclic components whose periods are
shorter than Duration input
alpha1 = (1 - Sine (360 / Duration)) / Cosine(360 /
Duration);
HP = .5*(1 + alpha1)*(Close - Close[1]) + alpha1*HP[1];
//Smooth with a Super Smoother Filter from equation 3-3
a1 = expvalue(-1.414*3.14159 / 10);
b1 = 2*a1*Cosine(1.414*180 / 10);
c2 = b1;
c3 = -a1*a1;
c1 = 1 - c2 - c3;
Filt = c1*(HP + HP[1]) / 2 + c2*Filt[1] + c3*Filt[2];
//3 Bar average of Wave amplitude and power
Wave = (Filt + Filt[1] + Filt[2]) / 3;
Pwr = (Filt*Filt + Filt[1]*Filt[1] + Filt[2]*Filt[2]) / 3;
//Normalize the Average Wave to Square Root of the Average
Power
Wave = Wave / SquareRoot(Pwr);

Plot1(Wave);
```

ภาพที่ 2.2 ขั้นตอนการคำนวณ EBSW [6]

2.6 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับ Reinforcement learning ในการลงทุนทั้งตลาด Crypto และ ตลาดอื่นๆที่ผู้วิจัยได้ศึกษาค้นคว้า สามารถสรุปได้ดังนี้

2.6.1 Robust Forex Trading With Deep Q Network

โดย Sutta Sornmayura (2017) [1] งานชิ้นนี้เป็นการศึกษาเกี่ยวกับการใช้ปัญญาประดิษฐ์ในการพัฒนาระบบการเทรด Forex ในสกุลเงิน EUR/USD และ USD/JPY โดยใช้ Deep Q Network (DQN) โดยมี State หรือ Input 7 ชนิด ได้แก่

(1) ราคาปิด

(2) ผลต่างของราคาวันนี้กับวันก่อน

(3) เส้นค่าเฉลี่ย 10 วัน (Sma 10)

(4) เส้นค่าเฉลี่ย 50 วัน (Sma 50)

(5) เส้นค่าเฉลี่ย 100 วัน (Sma 100)

(6) เส้นค่าเฉลี่ย 10 วัน - เส้นค่าเฉลี่ย 50 วัน

(7) Sine-wave indicators

ด้วยโครงสร้างแบบ MLP ที่ประกอบด้วย

(1) Input layers 7 nodes

(2) Hidden Layers 48 nodes

(3) Output Layers 4 nodes

ซึ่ง Output Layers จะเป็น Action ของปัญญาประดิษฐ์ คือ Buy,Sell,Close และ ไม่ทำอะไร แล้วได้ทำการเปรียบเทียบผลการทำงานของปัญญาประดิษฐ์กับกลยุทธ์ Buy-and-Hold และ เทรดเดอร์ผู้เชี่ยวชาญ

โดยมีข้อสมมติฐานคือ

(1) เงินลงทุนเริ่มต้น 100,000 US

(2) No Transaction Cost

(3) ซื้อ หรือ ขาย ครั้งละ 1% ของเงินต้น

(4) เปิด Position ได้ครั้งละ 1 Position

(5) ใช้ราคาปิดในการ Buy หรือ Sell

t-Test: Paired Two Sample for Means		
	Annual Returns_agent	Annual Returns_B&H
Mean	43.88866667	1.466
Variance	5056.348212	108.3599257
Observations	15	15
Pearson Correlation	0.477950253	
Hypothesized Mean Difference	0	
df	14	
t Stat	2.461020542	
P(T<=t) one-tail	0.013727607	
t Critical one-tail	1.761310136	
P(T<=t) two-tail	0.027455215	
t Critical two-tail	2.144786688	

ภาพที่ 2.3 การเปรียบเทียบผลลัพธ์ในคู่เงิน EUR/USD ระหว่าง ปัญญาประดิษฐ์ กับ กลยุทธ์ Buy and hold

ที่มา: https://core.ac.uk/download/pdf/233618241.pdf

t-Test: Paired Two Sample for Means		
	Annual Returns	Annual Returns_CTA
Mean	43.88866667	3.934666667
Variance	5056.348212	28.88141238
Observations	15	15
Pearson Correlation	-0.035775111	
Hypothesized Mean Difference	0	
df	14	
t Stat	2.164144073	
P(T<=t) one-tail	0.024114189	
t Critical one-tail	1.761310136	
P(T<=t) two-tail	0.048228379	
t Critical two-tail	2.144786688	

ภาพที่ 2.4 การเปรียบเทียบผลลัพธ์ในคู่เงิน EUR/USD ระหว่าง ปัญญาประดิษฐ์ กับ เทรดเดอร์ผู้เชี่ยวชาญ

ที่มา: https://core.ac.uk/download/pdf/233618241.pdf

t-Test: Paired Two Sample for Means		
	Annual Returns_agent	Annual Returns_B&H
Mean	26.732	0.925333333
Variance	2255.99946	142.8156552
Observations	15	15
Pearson Correlation	0.076078354	
Hypothesized Mean Difference	0	
df	14	
t Stat	2.078459449	
P(T<=t) one-tail	0.028269352	
t Critical one-tail	1.761310136	
P(T<=t) two-tail	0.056538704	
t Critical two-tail	2.144786688	

ภาพที่ 2.5 การเปรียบเทียบผลลัพธ์ในคู่เงิน USD/JPY ระหว่าง ปัญญาประดิษฐ์ กับ กลยุทธ์ Buy and hold

ที่มา: https://core.ac.uk/download/pdf/233618241.pdf

t-Test: Paired Two Sample for Means		
	Annual Returns_agent	Annual Returns_CTA
Mean	26.732	3.934666667
Variance	2255.99946	28.88141238
Observations	15	15
Pearson Correlation	-0.474885183	
Hypothesized Mean Difference	0	
df	14	
t Stat	1.756304525	
P(T<=t) one-tail	0.0504389	
t Critical one-tail	1.761310136	
P(T<=t) two-tail	0.1008778	
t Critical two-tail	2.144786688	

ภาพที่ 2.6 การเปรียบเทียบผลลัพธ์ในคู่เงิน USD/JPY ระหว่าง ปัญญาประดิษฐ์ กับ เทรดเดอร์ผู้เชี่ยวชาญ

ที่มา: https://core.ac.uk/download/pdf/233618241.pdf

โดยผลการวิจัยพบว่าตัวแทนปัญญาประดิษฐ์สามารถทำได้ได้ดีกว่ามนุษย์ ได้อย่างมีประสิทธิผลอย่างมีนัยยะสำคัญ

2.6.2 Dynamic portfolio rebalancing through reinforcement learning

โดย Lim, Q.Y.E., Cao, Q. & Quek (2021) [2] งานชิ้นนี้เป็นการศึกษาเกี่ยวกับการทดสอบการปรับสมดุลย์ 4 แบบ กับ สินทรัพย์ 3 ประเภท ด้วย Deep Q Network (DQN) โดยใน 1 สินทรัพย์จะประกอบด้วย สินทรัพย์เสี่ยงสูง,เสี่ยงกลาง,เสี่ยงน้อย สินทรัพย์ 3 ประเภท ได้แก่

(1) ดัชนีตลาดหลักทรัพย์ : บราซิล(BVSP),ไต้หวัน(TWII),อเมริกา(Nasdaq Composite(IXIC))

(2) หุ้นในตลาด S&P 500 : American Express(AXP), McDonald(MCD), Walmart(WMT)

(3) หุ้นในตลาด Nasdaq : UMB Financial (UMBF),Uniti Group (UNIT), Mandiant (MNDT)

โดย Agent จะ Action เมื่อมีการทำนายว่าเกิดการกลับตัวของแนวโน้มโดย Action มี 4 แบบ

(1) เพิ่มสัดส่วนสินทรัพย์เสี่ยงสูงตาม composition rate แล้วลดสินทรัพย์อื่น

(2) เพิ่มสัดส่วนสินทรัพย์เสี่ยงกลางตาม composition rate แล้วลดสินทรัพย์อื่น

(3) เพิ่มสัดส่วนสินทรัพย์เสี่ยงสูง และ กลาง ตาม composition rate แล้วลดสินทรัพย์อื่น

(4) เพิ่มแต่สินทรัพย์เสี่ยงต่ำ ตาม composition rate แล้วลดสินทรัพย์อื่น

ซึ่ง composition rate คือ

$$\text{Composition rate} = 1 - (\text{BCR} * (\text{n} - 1))$$

เมื่อ BCR คือ Base Composition Rate ซึ่งมีค่าเท่ากับ 0.1

n คือ เท่ากับจำนวนสินทรัพย์

ซึ่งสัดส่วนของสินทรัพย์จะอยู่ระหว่าง Composition rate และ Base Composition Rate ซึ่งการปรับสมดุลย์พอร์ตการลงทุนแบบเต็มจำนวน (Full Portfolio Rebalancing) สัดส่วนจะเปลี่ยนไปมาระหว่าง 2 ค่า คือ 0.1 และ 0.8 ซึ่งพบว่าเป็นการเปลี่ยนแปลงที่มากเกินไปจึงมีการใช้การทยอยปรับสมดุลย์พอร์ตการลงทุน (Gradual Portfolio Rebalancing) ครั้งละ k% โดยค่า k ค่าที่เหมาะสมคือ 0.3 State ประกอบด้วย

(1) Standardized 6 วัน ของ EMA 15 วัน ของ สินทรัพย์เสี่ยงสูง

(2) Standardized 6 วัน ของ EMA 15 วัน ของ สินทรัพย์เสี่ยงกลาง

(3) Standardized 6 วัน ของ ส่วนต่างของ MACD กับ Signal line 15 วัน ของสินทรัพย์เสี่ยงสูง

(4) Standardized 6 วัน ของ MACD 15 วัน ของ สินทรัพย์เสี่ยงกลาง

(5) จำนวนวันตั้งแต่วันที่เกิด การเปลี่ยนแนวโน้มครั้งล่าสุด จนถึงวันปัจจุบัน

การปรับสมดุลย์ 4 แบบ คือ

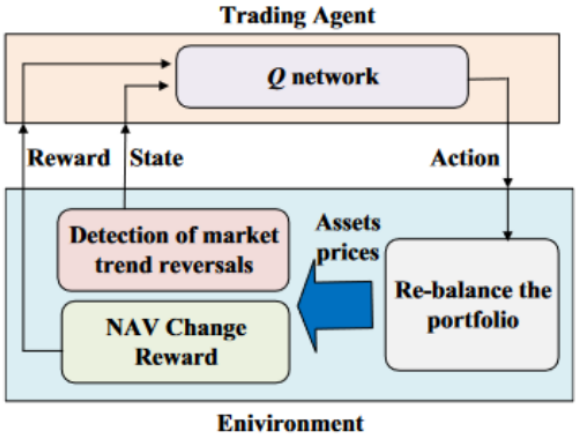
(1) การปรับสมดุลย์พอร์ตการลงทุนแบบเต็มจำนวนแบบไม่มีโมเดลพยากรณ์

(2) การปรับสมดุลย์พอร์ตการลงทุนแบบเต็มจำนวนแบบมีโมเดลพยากรณ์

(3) การทยอยปรับสมดุลย์พอร์ตการลงทุนแบบไม่มีโมเดลพยากรณ์

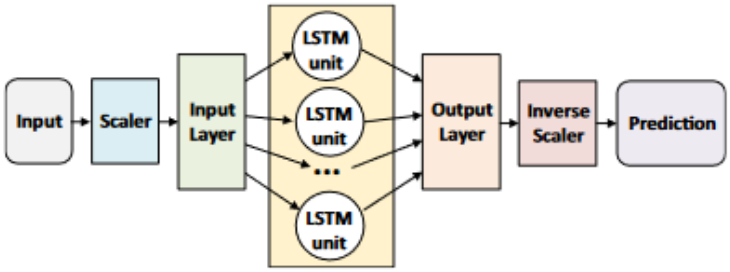
(4) การทยอยปรับสมดุลย์พอร์ตการลงทุนมีโมเดลพยากรณ์

ซึ่ง Agent จะ Action ทุกครั้งที่ โมเดลพยากรณ์ทำนายว่าจะมีการเปลี่ยนแปลงในแนวโน้มเกิดขึ้น โดยโครงสร้างของโมเดลพยากรณ์จะใช้ LSTM สำหรับรูปแบบที่ไม่มีโมเดลพยากรณ์จะใช้ Technical Indicator : MACD แทน



ภาพที่ 2.7 โครงสร้างของ Deep Q Network

ที่มา : https://doi.org/10.1007/s00521-021-06853-3



ภาพที่ 2.8 โครงสร้างโมเดลพยากรณ์แบบ LSTM

ที่มา : https://doi.org/10.1007/s00521-021-06853-3

	(1) Full rebalance without prediction (%)	(2) Full rebalance with prediction (%)	(3) Gradual rebalance without prediction (%)	(4) Gradual rebalance with prediction (%)
NAV Return % by end of 2018				
RL agent	27.3	39.7	47.2	55.2
BVSP	74.3			
TWII	12.7			
IXIC	48.7			
NAV max drop %				
RL agent	−23.7	−6.8	−6.5	0.7
BVSP	−25.3			
TWII	−14.3			
IXIC	−4.7			

ภาพที่ 2.9 ผลการทดลองกับดัชนีตลาดหลักทรัพย์

ที่มา : https://doi.org/10.1007/s00521-021-06853-3

	(1) Full rebalance without prediction (%)	(2) Full rebalance with prediction (%)	(3) Gradual rebalance without prediction (%)	(4) Gradual rebalance with prediction (%)
NAV Return % by end of 2018				
RL agent	21.7	29.3	36.7	63.3
AXP	51.6			
MCD	51.7			
WMT	40.7			
NAV max drop %				
RL agent	−14.6	−6.7	−6.6	−6.9
AXP	−6.3			
MCD	−1.7			
WMT	−17.5			

ภาพที่ 2.10 ผลการทดลองกับหุ้นในตลาด S&P 500

ที่มา : https://doi.org/10.1007/s00521-021-06853-3

	(1) Full rebalance without prediction (%)	(2) Full rebalance with prediction (%)	(3) Gradual rebalance without prediction (%)	(4) Gradual rebalance with prediction (%)
NAV Return % by end of 2018				
RL agent	−46.7	−20.7	−8.7	46.7
UMBF	34.3			
UNIT	−20.6			
MNDT	−25.7			
NAV max drop %				
RL agent	−49.3	−25.3	−22.7	−20.5
UMBF	−9.7			
UNIT	−26.7			
MNDT	−52.3			

ภาพที่ 2.11 ผลการทดลองกับดัชนีตลาดหลักทรัพย์

ที่มา : https://doi.org/10.1007/s00521-021-06853-3

โดยผลการศึกษาพบว่าการปรับสมดุลย์ ในแบบการทยอยปรับสมดุลย์พอร์ตการลงทุนแบบมีโมเดลพยากรณ์ ให้ผลตอบแทนมากที่สุด และ Max Drop % ที่น้อยที่สุดในทุกๆสินทรัพย์

2.6 .3 Multi-Agent Deep Reinforcement Learning With Progressive Negative Reward for Cryptocurrency Trading โดย Kittiwin Kumlungmak และ Peerapon Vateekul (2023) [3] ได้นำเสนอเกี่ยวกับการใช้ Multi-Agent Proximal Policy Optimization ใน Cryptocurrency 6 ชนิด คือ

(1) Cardano (ADAUSDT)

(2) Binance Coin (BNBUSDT)

(3) Bitcoin (BTCUSDT)

(4) Ethereum (ETHUSDT)

(5) Ripple (XRPUSDT)

(6) USDT

โดย State จะประกอบด้วย 2 ส่วน คือ

(1) Portfolio Data คือ สัดส่วนของ Cryptocurrency ทั้ง 6 ตัว

(2) Market Data คือ ราคาปิด, MACD, RSI, CCI, ADX 5 วันย้อนหลัง

Action จะประกอบด้วย Buy, Sell, Hold โดยให้ Agent เลือกค่าระหว่าง -1 ถึง 1 โดยค่าติดลบคือ Sell ค่า-บวกคือการ Buy ซึ่งจากนั้นจะนำไป คูณด้วย 100,000 เพื่อให้เป็นมูลค่าที่ต้องซื้อ

สำหรับ Reward จะประกอบไปด้วย 2 ส่วนคือ Local Reward และ Global Reward โดย Local Reward จะใช้กับการคำนวณ Reward เฉพาะใน เหรียญนั้นๆ Global Reward จะใช้กับการคำนวณ

Agent ในทุกๆตัว โดยงานวิจัยชิ้นนี้ได้เน้นการนำเสนอ Multi-Scale Continuous Loss Reward (MSCL) ซึ่งคือ

$$r_{MSCL}(S_t, A_t, S_{t+1}) = \mu \cdot |n_{loss} - n_{win}| \cdot r_{portfolio,t}$$

โดยให้ μ >= 0 โดยมีจุดประสงค์เพื่อให้ Agent ลดการทำผิดพลาดในสิ่งเดิมๆซ้ำผ่านการเพิ่มค่าติดลบใน Reward โดยทีเงื่อนไข คือ

(1) มูลค่า มูลค่าสินทรัพย์สุทธิขณะนั้น น้อยกว่า มูลค่าสินทรัพย์สุทธิตอนเริ่มต้น

(2) จำนวนครั้งของการเทรดขาดทุนมากกว่ากำไร

(3) รางวัลจากการเปลี่ยนแปลงมูลค่าของพอร์ตการลงทุน <= 0

Method	Cumulative Return (percent)	Sharpe Ratio	Calmar Ratio	Volatility (percent)	MDD (percent)
Overall Test Set					
LSTM	-44.29	-0.85	-0.63	89.41	-70.25
GRU	-54.13	-1.53	-0.84	79.93	-64.08
ResNet	-57.88	-1.59	-0.86	84.90	-67.67
Res2Net	-37.91	-0.62	-0.59	89.12	-64.68
MLP	-17.47	-0.12	-0.35	75.69	-49.81
Bullish Test Set					
LSTM	52.61	5.9	2.97	91.63	-17.71
GRU	31.36	4.98	2.36	69.44	-13.31
ResNet	23.7	4.03	1.75	67.96	-13.57
Res2Net	34.18	4.55	2.00	83.93	-17.06
MLP	28.26	4.64	2.35	68.25	-12.02
Bearish Test Set					
LSTM	-3.86	-1.99	-0.47	22.74	-8.23
GRU	-25.56	-6.25	-0.98	55.08	-26.02
ResNet	-29.56	-6.8	-0.96	60.08	-30.89
Res2Net	-11.21	-2.92	-0.71	45.99	-15.87
MLP	2.36	2.83	1.14	10.26	-2.07
Up-Down Test Set					
LSTM	1.59	1.47	0.38	13.22	-4.18
GRU	1.43	0.58	0.08	53.14	-17.24
ResNet	-5.61	-0.84	-0.23	59.83	-24.63
Res2Net	1.15	0.6	0.12	30.14	-9.31
MLP	2.35	2.76	1.13	10.09	-2.07
Sideways test set					
LSTM	-3.25	-0.11	-0.20	78.28	-16.01
GRU	-6.66	-0.55	-0.27	83.94	-24.44
ResNet	-4.01	-0.19	-0.21	80.94	-19.36
Res2Net	6.52	1.22	0.29	110.9	-22.19
MLP	1.21	0.62	0.07	93.29	-17.59

ภาพที่ 2.15 ผลการทดลองกับโครงสร้างต่างๆ [3]

ซึ่งจากผลการทดลองโครงสร้างแบบ MLP ทำได้ดีจาก 3 ใน 5 Dataset และ ผลของ MSCL Reward เมื่อนำไปใช้รวมกับ Reward ประเภทอื่นๆผลของ MSCL จะยังคงเด่นชัดอยู่ส่งผลให้ Cumulative Returns, Shrape Ratio, Calmar Ratio มีค่าออกมาค่อนข้างดี

บทที่ 3
ระเบียบวิธีวิจัย

การประยุกต์ใช้ Reinforcement learning ด้วยวิธีการ Double Deep Q-learning มาใช้ในการปรับสมดุลย์ Bitcoin และ การขายชอร์ตสัญญาซื้อขายล่วงหน้า ใน Multi-Asset Mode มีขั้นตอนและรายละเอียด ดังต่อไปนี้

3.1 การจำลอง Environment ของ Cryptocurrency Exchange

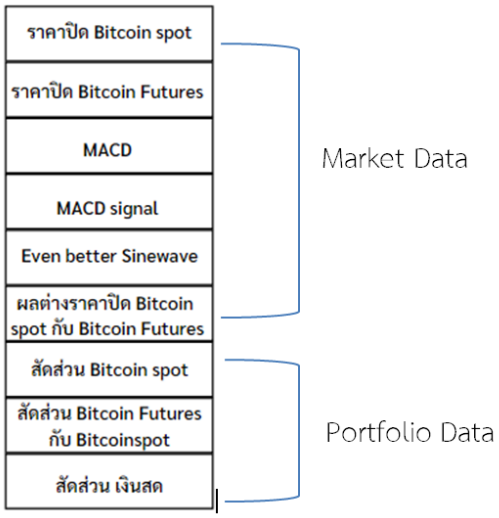
เพื่อให้ Agent สามารถเรียนรู้เพื่อเทรดในสินค้า Bitcoin และ การขายชอร์ตสัญญาซื้อขายล่วงหน้า ได้ Environment ต้องมีส่วนประกอบดังนี้

3.1.1 State

ใช้ข้อมูล Bitcoin spot ,สัญญาซื้อขายล่วงหน้าของ Bitcoin และ funding rate ของ BTC/USDT จากเว็บ Binance โดยแบ่งเป็น Train Dataset ใช้ตั้งแต่ 19/7/2020-23/6/2022 สำหรับ Test Dataset ใช้ตั้งแต่ 24/6/2022 - 06/06/2023 ซึ่ง State ประกอบด้วย 2 ส่วน [3] คือ

(1) Market Data ราคาปิด Bitcoin spot, ราคาปิดสัญญาซื้อขายล่วงหน้าของ Bitcoin, MACD , MACD signal , Even better Sinewave [1] , ผลต่างราคาปิด Bitcoin spot กับ สัญญาซื้อขายล่วงหน้าของ Bitcoin

(2) Portfolio Data คือ สัดส่วนมูลค่าของ Bitcoin spot ต่อมูลค่า พอร์ตการลงทุน, สัดส่วนเงินสด ต่อมูลค่า พอร์ตการลงทุน, สัดส่วนสัญญาซื้อขายล่วงหน้าของ Bitcoin ต่อ จำนวน Bitcoin spot

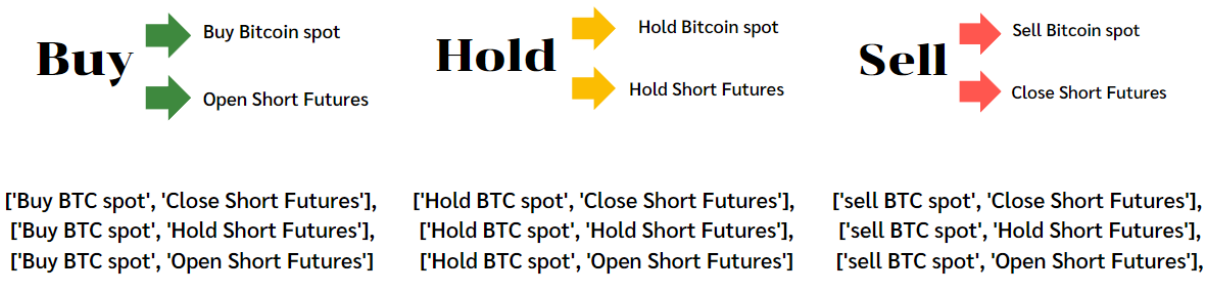


ภาพที่ 3.1 State หรือ Features Input 9 Features

โดยจากรูปที่ 3.1 Technical Indicator ทั้ง 2 ชนิด ได้แก่ MACD, Even better Sinewave จะเรียกใช้ผ่าน Python Library ที่ชื่อ Pandas TA

3.1.2 Action

Action มี 3 แบบ คือ Buy , Hold , Sell ใน 1 Asset ซึ่งจะกระทำกับ Bitcoin spot และ การขายชอร์ตสัญญาซื้อขายล่วงหน้าของ Bitcoin พร้อมกัน ดังนั้น Action ที่เกิดขึ้นได้ ทั้งหมด 9 Actions ซึ่งจะทำการ Action ณ ราคาปิดของวันนั้นๆ ดังรูปที่ 3.2



ภาพที่ 3.1 แสดง Action ทั้งหมด 9 รูปแบบ

3.1.3 Reward

Agent จะเรียนรู้ผ่าน Reward จากการ Action ในแต่ละ State โดย Reward ที่ใช้ในงานสารนิพนธ์นี้ ประกอบด้วย 2 ส่วน

1. NAV reward คือ มูลค่าสินทรัพย์ใน พอร์ตการลงทุน (Net Asset Value) ณ ช่วงเวลานั้น

เทียบกับช่วงก่อน โดยคำนวณจาก

$$\text{NAV reward} = \frac{\text{NAV}_{t+1}}{\text{NAV}_t} - 1$$

เมื่อ NAV_{t+1} คือ มูลค่าสินทรัพย์สุทธิของ State ถัดไปหลังจากมีการ Action ไปแล้ว

NAV_t คือ มูลค่าสินทรัพย์สุทธิของ State ก่อนมีการ Action

2. Multi-scale continuous loss (MSCL) Reward เพื่อลดโอกาสการลดลงอย่างต่อเนื่องของมูลค่าสินทรัพย์สุทธิ ซึ่งมีแนวคิดมาจากการเพิ่มบทลงโทษที่มากขึ้นในแต่ละครั้ง ถ้าทำผิดแบบเดิมๆซ้ำๆ ดังนั้นหากเกิดเงื่อนไขเหล่านี้ขึ้นพร้อมกัน Reward จะติดลบมากขึ้น โดยคำนวณจากสมการด้านล่าง [3]

$$\text{MSCL reward} = \text{NAV reward}_t + \left[\left(\frac{\text{NAV}_t}{\text{NAV}_{start}} - 1\right) \cdot (\text{N}_{win} - \text{N}_{loss}) \cdot (\text{NAV reward}_t)\right]$$

เมื่อ N_{win} คือ จำนวนครั้งที่ชนะ

N_{loss} คือ จำนวนครั้งที่แพ้

โดย MSCL Reward จะถูกใช้คำนวณเมื่อเข้าเงื่อนไขทั้ง 3 ข้อดังนี้

(1) มูลค่า มูลค่าสินทรัพย์สุทธิ ณ ขณะนั้น น้อยกว่า มูลค่าสินทรัพย์สุทธิ ที่เริ่มต้น

(2) จำนวนครั้งการเทรดขาดทุนมากกว่ากำไร

(3) NAV Reward ณ Timestep นั้น น้อยกว่าเท่ากับ 0

3.1.4 Constraints

ขีดจำกัดหรือเงื่อนไขของการทดลองของสารนิพนธ์เล่มนี้ คือ

(1) ใช้ base composition rates (BCR) และ composition rates (CR) ในการควบคุมสัดส่วนขั้นต่ำและสัดส่วนสุง (ขีดจำกัด) สุดของ Bitcoin sport กับ เงินสด เท่านั้น เพื่อป้องกันไม่ให้มี Bitcoin หรือเงินสด 100% ใน พอร์ตการลงทุน จาก สมการด้านล่างนี้

$$\text{Composition rate} = 1 - (\text{BCR} * (\text{n} - 1))$$

โดย n มีค่าเท่ากับ 2 (1.Bitcoin และ 2.เงินสด) BCR มีค่าเท่ากับ 0.1และ CR จึงมีค่าเท่ากับ 0.9

(2) ในกรณีของ Bitcoin ปริมาณในการซื้อขายแต่ละครั้ง (k) ให้มีค่า เท่ากับ 30% ของมูลค่าพอร์ตการลงทุน สำหรับ การขายชอร์ตสัญญาซื้อขายล่วงหน้า จะให้ซื้อแต่ละครั้ง (k) เท่ากับ 30% ของ Bitcoin ที่มีอยู่ ส่วนการขายจะให้ขายแต่ละครั้ง (k) เท่ากับ 30% หรือ ขายออกไปได้เท่าที่มีเหลืออยู่

(3) ใช้ Multi Asset Mode ซึ่งทาง Binance เปิดให้สามารถนำเอาทรัพย์สินใน พอร์ตการลงทุน ไปวางเป็นหลักประกันเพื่อเปิดสัญญา สัญญาซื้อขายล่วงหน้า เพื่อ ป้องกันความเสี่ยง ได้โดยคิดมูลค่าเท่ากับ 95% ของมูลค่าทรัพย์สินที่นำมาวาง และ เพื่อป้องกันการถูก Call Margin จะไม่ให้ เปิดขายชอร์ต

สัญญาซื้อขายล่วงหน้า มากกว่าจำนวน Bitcoin ที่มี หรือ Leverage เท่ากับ 0 ซึ่งต้องมี Bitcoin spot ก่อนเสมอ จึงสามารถ เปิดขายชอร์ตสัญญาซื้อขายล่วงหน้า ได้ ดังนั้น ปริมาณ การขายชอร์ตสัญญาซื้อขายล่วงหน้า ที่มีสุงสุด ณ Time step นึงจะเท่ากับ

$$\text{Max Short Futures Position} = 0.95 \cdot \text{Bitcoin spot}$$

(4) เมื่อมีการปิดสัญญา การขายชอร์ตสัญญาซื้อขายล่วงหน้า แล้วขาดทุนก็จะทำการขาย Bitcoin spot ชดเชยผลขาดทุนนั้นเพื่อไม่ให้เงินสดติดลบเนื่องจาก การขายชอร์ตสัญญาซื้อขายล่วงหน้า ทำการชำระราคาด้วย USDT ซึ่งจะนับเป็นการเทรดแบบ loss และ แต่จะไม่ใช้กรณีนี้กับ Bitcoin spot เพราะ ในการซื้อ Spot แต่ละครั้งจะไม่คำนึงถึง การขายชอร์ตสัญญาซื้อขายล่วงหน้า และ ใช้ประเภทบัญชีในการซื้อขาย Spot ซึ่งเงินไม่ติดลบ

(5) เริ่มต้นด้วยการ พอร์ตการลงทุน เงินสด 50% , Bitcoin 50%, ป้องกันความเสี่ยง 0.95 ของปริมาณ Bitcoin Spot ที่อยู่ใน พอร์ตการลงทุน

(6) ค่าคอมการขายชอร์ตสัญญาซื้อขายล่วงหน้า ใช้ Taker Rate 0.04% ส่วน Spot ใช้ 0.10%

(7) การถือ การขายชอร์ตสัญญาซื้อขายล่วงหน้า จะถูกคิด Funding Rate เนื่องจากสัญญา การขายชอร์ตสัญญาซื้อขายล่วงหน้า ที่ใช้เป็นแบบที่ไม่หมดอายุ (Perpetual Futures)ซึ่งราคา สัญญาซื้อขายล่วงหน้า ในตลาดสามารถมากกว่า หรือ น้อยกว่า Bitcoin Spot ได้ Funding Rate จึงเป็นตัวปรับให้ราคาสัญญาซื้อขายล่วงหน้า ยังอยู่บนพื้นฐานราคาจริง

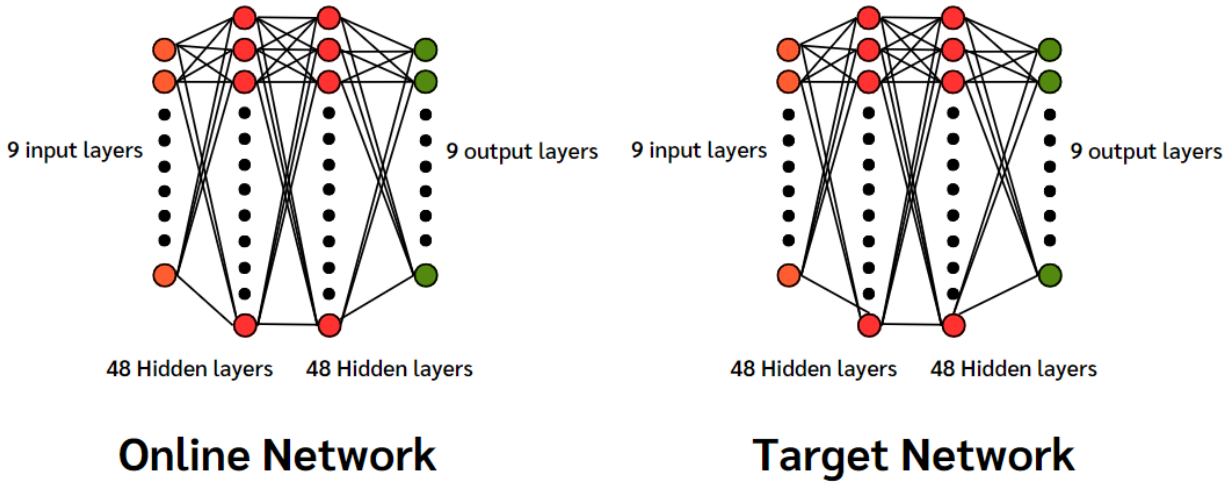
7.1 หากราคา สัญญาซื้อขายล่วงหน้า ต่ำกว่าราคา Spot หรือเรียกว่า Discount (Funding rate negative) ผู้ที่เปิด Position short ต้องจ่ายให้ผู้เปิด position long

7.2 หากราคา สัญญาซื้อขายล่วงหน้า สูงกว่าตลาด Spot หรือเรียกว่า Premium (Funding rate positive) ผู้ที่เปิด Position long ต้องจ่ายให้ผู้ที่เปิด position short

(8) ระบบนับกำไรขาดทุนแบบ FIFO

3.1.5 Algorithm และ โครงสร้างของ Network

ในงานสารนิพนธ์ใช้ Algorithm ที่ชื่อว่า Double Deep Q Network ซึ่งช่วยลดปัญหา Overestimation ใน Deep Q Network [4] โดยมีโครงสร้าง ประกอบด้วย



ภาพที่ 3.2 ภาพแสดงโครงสร้างของ Network .o Double Deep Q Network

(1) Online Network

(2) Target Network

ซึ่งจากรูปที่ 3.3 โครงสร้างของ Network ทั้ง 2 จะใช้ Input layer ใช้ 9 Nodes ซึ่งแทนแต่ละ Feature ใน State, Hidden layer 2 layer ซึ่งใน 1 layer จะใช้ 48 Nodes ,Output layer ใช้ 9 Nodes ซึ่งแทน Action ที่ Agent สามารถเลือกได้ โดยในสารนิพนธ์ชิ้นนี้จะมีการใช้

(1) Experience Replay ใช้การเก็บข้อมูลแบบ deque ซึ่งเก็บข้อมูลได้ 20,000 ชุด เพื่อสุ่มหยิบขึ้นมาเทรน Neural Network

(2) Soft update ค่า Weight จาก Online Network จะถูก Update ค่าใน Target Network ทุกๆ Timestep ผ่านสมการดังนี้ โดยให้ τ มีค่าเท่ากับ 0.0001

$$\theta' = \tau\theta + (1-\tau)\theta'$$

เมื่อ θ คือ ค่า Weight ของ Online Network

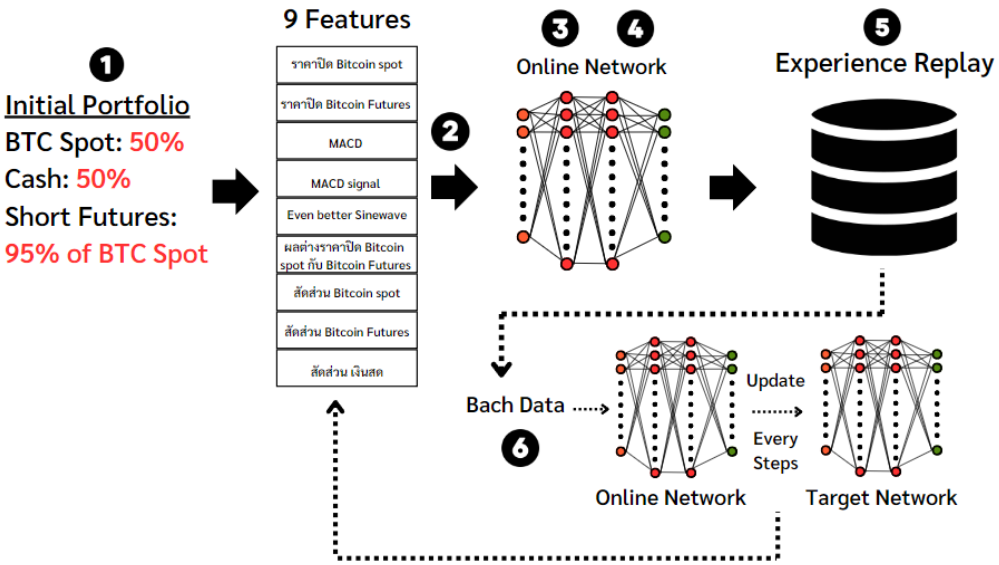
θ' คือ ค่า Weight ของ Target Network

3.2 การเตรียมข้อมูล

ใช้ข้อมูล Bitcoin spot ,สัญญาซื้อขายล่วงหน้า Bicoin และ Funding Rate ของ BTC/USDT จากเว็บ Binance โดย 40 วันแรกจะถูกสร้างเป็น Technical Indicator ใน State แรกโดยแบ่งเป้น Train Dataset ใช้ตั้งแต่ 19/7/2020-23/6/2022 สำหรับ Test Dataset ใช้ตั้งแต่ 24/6/2022 -06/06/2023 โดยใช้ Min-Max Normalization

3.3 การ Training

ทำการ Train ผ่าน Google Colab โดยตั้งค่า Hyper-parameters ต่างๆ ดังนี้



ภาพที่ 3.4 ภาพแสดงขั้นตอนการแสดงการ Train ของ Double Deep Q Network

(1) Alpha = 0.00001 (2) Gamma = 0.95 (3) Epsilon Decay = 0.999 (4) Batch Size = 32 (5) เงินลงทุนเริ่มต้น เท่ากับ 10,000 $

(1) จะเริ่มต้นด้วยการ เริ่มต้นพอร์ตการลงทุน ขึ้นมาด้วยอัตราส่วน Bitcoin 50%, เงินสด 50% และ ขายชอร์ตสัญญาซื้อขายล่วงหน้า 95% ของ Bitcoin ให้ Epsilon = 1.00

(2) จากนั้นก็จะเปลี่ยนเป็น State ถัดไปโดยนำเอา State เป็นข้อมูลเพื่อ Input เข้าสู่ Online Network

(3) หลังจาก Online Network เลือก Action ได้แล้ว ก็จะเก็บ มูลค่าสินทรัพย์สุทธิ ปัจจุบันเป็น มูลค่าสินทรัพย์สุทธิ ใน Timestep ก่อนหน้า แล้วทำการเปลี่ยนเป็น State ถัดไป

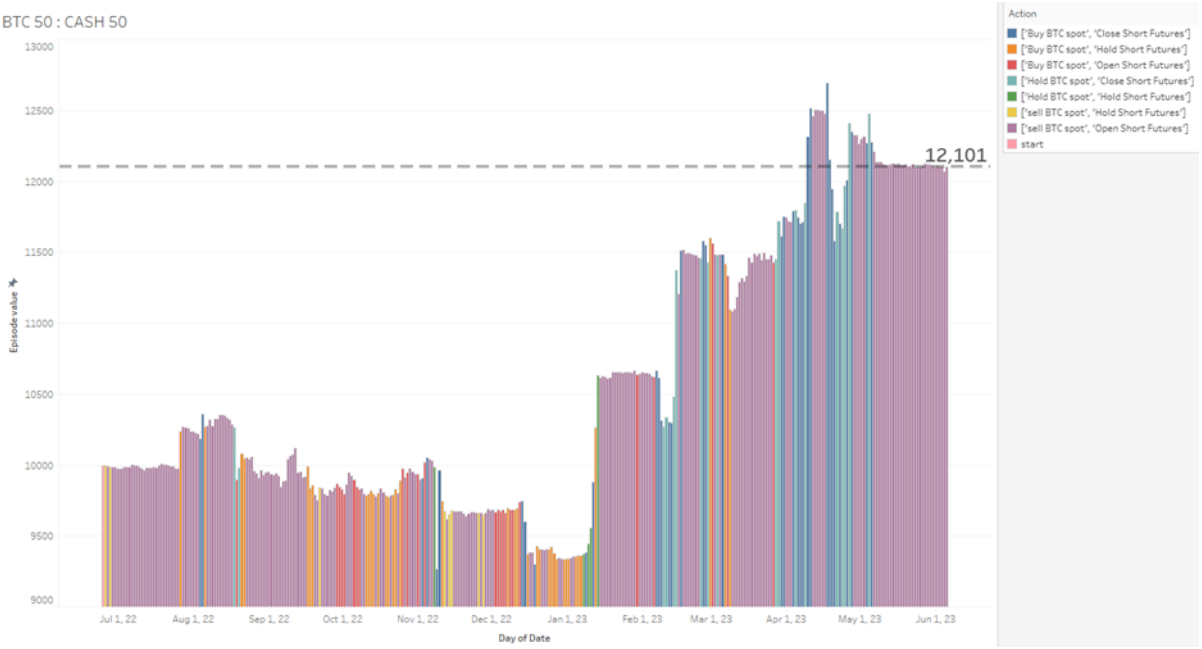
(4) คำนวณ มูลค่าสินทรัพย์สุทธิ ใหม่อีกครั้ง โดยกำหนดให้เป็น มูลค่าสินทรัพย์สุทธิ ใน Timestep ปัจจุบัน แล้วคำนวณ Reward ให้กับ Agent

(5) เก็บข้อมูล State, Action, Reward,State ถัดไป ลงใน Experience Replay

(6) สุ่ม Experience Replay มาตามขนาด Batch size เพื่อ Train Neural Network และ Update Target แล้วลดค่า Epsilon ด้วย การคูณกับ Epsilon Decay จากนั้นก็จะเริ่มใหม่ในข้อ 2 โดยในสารนิพนธ์ชิ้นนี้ได้ทำการ Train ไป 1800 Episodes

บทที่ 4
ผลการวิจัย

สารนิพนธ์เล่มนี้ได้ประยุกต์การใช้ Double Deep Q Network ในการปรับสมดุลย์ สินทรัพย์ 2 ชนิด คือ Bitcoin Spot และ เงินสด (USDT) และ มีการป้องกันความเสี่ยงด้วยการ การขายชอร์ตสัญญาซื้อขายล่วงหน้าโดยใช้ข้อมูล Bitcoin spot , สัญญาซื้อขายล่วงหน้า Bitcoin และ Funding rate ของ BTC/USDT จากเว็บ Binance สำหรับการTrain ใช้ตั้งแต่ 19/7/2020 - 23/6/2022 สำหรับการTest ใช้ตั้งแต่ 24/6/2022 - 06/06/2023



ภาพที่ 4.1 ภาพแสดง มูลค่าสินทรัพย์สุทธิ ของ พอร์ตการลงทุนและการเลือก Action ของ Agent

จากรูป 4.1 แสดงผลผ่านโปรแกรม Tableau โดยแกนตั้งแสดงถึง มูลค่าสินทรัพย์สุทธิ ของพอร์ตการลงทุน แกนนอนคือวันที่ปรับสมดุลย์ในTest Dataset จะเห็นได้ว่า มูลค่าสินทรัพย์สุทธิ ของ พอร์ตการลงทุน ณ Terminal State อยู่ที่ 12,100.96 $ และ มี Action ที่ถูกเลือกทั้งหมด 7 Action จาก 9 Action จากนั้นจะนำผลจากการ Test มาวัดผลผ่าน ตัววัดผลการดำเนินงานของพอร์ตการลงทุน(Portfolio Performance Metrics) 5 ชนิด

4.1 ตัววัดผลการดำเนินงานของพอร์ตการลงทุน

4.1.1 Cumulative Returns

คือ ผลตอบแทนโดยรวมของ พอร์ตการลงทุน ซึ่งจะดูผลกระทบจากการเปลี่ยนแปลงของราคาต่อมูลค่าเงินลงทุนทั้งหมด พอร์ตการลงทุน โดยมีสมการคือ

$$\textbf{Cumulative Returns} = \frac{\mathbf{V_{end}} - \mathbf{V_{start}}}{\mathbf{V_{start}}}$$

เมื่อ $\mathbf{V_{end}}$ = NAV ของ Portfolio ณ State สุดท้าย

$\mathbf{V_{start}}$ = NAV ของ Portfolio ณ State เริ่มต้น ซึ่ง เท่ากับ 10,0000$

จากผลการทดลองได้ค่า Cumulative Returns เท่ากับ 21.00965%

4.1.2 Annualized Volatility

คือ การวัดความผันผวนของ พอร์ตการลงทุน ซึ่งสะท้อนถึงความเสี่ยงหากมีค่าที่มากเกินไป โดยมีสมการดังนี้

$$\textbf{Annualized Volatility} = \textbf{Standard Deviation} * \sqrt{\textbf{Trading day}}$$

เมื่อ Standard Deviation = ค่า Standard Deviation ของผลตอบแทนในแต่ละวันจาก พอร์ตการลงทุนTrading day = จำนวนวันที่มีการเทรด

```
1 df_result
```

	Date	Action	Spot	Futures	Cash	Episode_value	daily_returns
0	2022-06-24	start	0.235430	0.223659	5000.000000	10000.000000	0.000000
1	2022-06-25	[sell BTC spot, Hold Short Futures]	0.235431	0.223659	4998.913147	9997.117989	-0.000288
2	2022-06-26	[sell BTC spot, Open Short Futures]	0.235431	0.223659	4997.941012	9993.470187	-0.000365
3	2022-06-27	[sell BTC spot, Hold Short Futures]	0.235431	0.223659	4997.131720	9987.260119	-0.000621
4	2022-06-28	[sell BTC spot, Open Short Futures]	0.235431	0.223659	4997.393017	9985.752155	-0.000151
...	...	...	...	...	...	...	...
342	2023-06-01	[sell BTC spot, Open Short Futures]	0.217847	0.196062	5961.614725	12107.656945	0.000705
343	2023-06-02	[sell BTC spot, Open Short Futures]	0.217847	0.196062	5960.334484	12102.507744	-0.000425
344	2023-06-03	[sell BTC spot, Open Short Futures]	0.217847	0.196062	5959.115205	12101.994288	-0.000042
345	2023-06-04	[sell BTC spot, Open Short Futures]	0.217847	0.196062	5957.794736	12070.593522	-0.002595
346	2023-06-05	[sell BTC spot, Open Short Futures]	0.217847	0.196062	5956.413618	12100.965766	0.002516

347 rows × 7 columns

ภาพที่ 4.2 แสดงผลการทดลองที่ได้จากการ Test ใน Google Colab

จากรูป 4.2 Trading Day มีจำนวนทั้งหมด 347 วัน ซึ่งเมื่อนำไปคูณกับค่า Standard Deviation ของผลตอบแทนในแต่ละวัน จะได้ค่า Annualized Volatility เท่ากับ 20.3568% ซึ่งมีค่าน้อยกว่าเมื่อเทียบกับความผันผวนของ Bitcoin ที่มีค่า 51.57076%

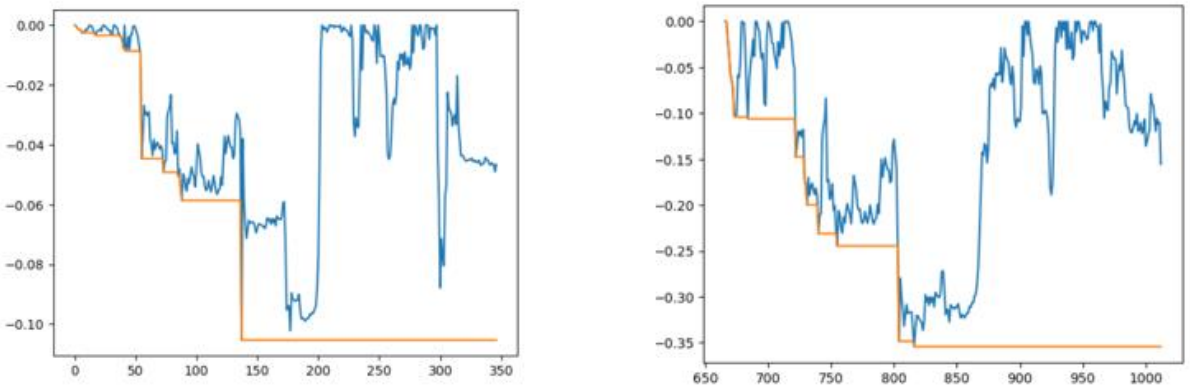
4.1.3 Max DrawDown

เป็นการวัดค่าผลขาดทุนสูงสุดที่เกิดขึ้นในช่วงที่ทำการทดสอบโดยมีสมการคือ

$$\textbf{Max Drawdown} = \frac{\mathbf{V_{trough}} - \mathbf{V_{peak}}}{\mathbf{V_{peak}}}$$

เมื่อ $\mathbf{V_{trough}}$ = มูลค่าสินทรัพย์สุทธิของพอร์ตการลงทุน ที่ต่ำที่สุด

$\mathbf{V_{peak}}$ = มูลค่าสินทรัพย์สุทธิของพอร์ตการลงทุน ที่สูงสุดก่อนเกิดจุดที่ต่ำที่สุด



ภาพที่ 4.3 แสดงผลของ Max Drawdown มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน เทียบกับ Bitcoin (ตามลำดับ) เส้นสีส้มคือ Max Drawdown ส่วนเส้นสีฟ้า คือ Drawdown

จากผลการทดสอบ มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน มีค่า Max Drawdown เท่ากับ 10.53967% ซึ่งต่ำกว่าเมื่อเทียบกับ Bitcoin ที่มีค่าเท่ากับ 35.43208%

4.1.4 Sharpe Ratio

คือ อัตราผลตอบแทนส่วนเกินต่อความผันผวน ซึ่งหากสัดส่วนยิ่งสูงนั่นหมายถึงผลตอบแทนยิ่งคุ้มค่าเมื่อเทียบกับความผันผวน โดยมีสมการดังนี้

$$\textbf{Sharpe ratio} = \frac{\mathbf{R_d - R_f}}{\boldsymbol{\sigma_d}}$$

เมื่อ $\mathbf{R_d}$ = ผลตอบแทนรายวันของ พอร์ตการลงทุน

$\mathbf{R_f}$ = ผลตอบแทนที่ปราศจากความเสี่ยง

$\boldsymbol{\sigma_d}$ = ความผันผวนของผลตอบแทนรายวันของ พอร์ตการลงทุน

โดย Risk free rate ใช้ค่าเฉลี่ยจาก Government bond 10 yrs. จาก สมาคมตราสารหนี้แห่งประเทศไทย แต่เนื่องจากการซื้อขายตราสารหนี้ หยุดวันเสาร์-อาทิตย์และนักขัตฤกษ์ ซึ่งมีข้อมูล 230 วัน จากนั้น เมื่อได้ค่าเฉลี่ยรายวันแล้ว จึงหาร 347 เพื่อหา Daily Returns ในช่วงที่ทำการทดลอง

Risk free rate ได้ค่าเท่ากับ 0.00768 % Standard Deviation ของผลตอบแทนรายวัน เท่ากับ 1.09281% ผลตอบแทนรายวันได้ 0.06086% จึงได้ค่า Sharpe Ratio 0.04865 เท่า แต่อยู่ในผลตอบแทนรายวัน จึงทำการ Annualize จึงได้ค่าเท่ากับ 0.90624 เท่า ซึ่งมีค่าน้อยกว่า 1 แสดงถึงว่า ผลตอบแทนอาจจะไม่คุ้มค่ากับความผันผวนมากนัก

4.1.5 Calmar Ratio

คือ อัตราผลตอบแทนส่วนเกินเทียบกับ Max Drawdown ซึ่งหากสัดส่วนยิ่งสูงนั่นหมายถึงผลตอบแทนยิ่งคุ้มค่าเมื่อ Max Drawdown

$$\textbf{Calmar ratio} = \frac{\mathbf{R_d - R_f}}{\mathbf{MDD}}$$

เมื่อ $\mathbf{R_d}$ = ผลตอบแทนรายวันของ พอร์ตการลงทุน

$\mathbf{R_f}$ = ผลตอบแทนที่ปราศจากความเสี่ยง

$\mathbf{MDD}$ = Max Drawdown

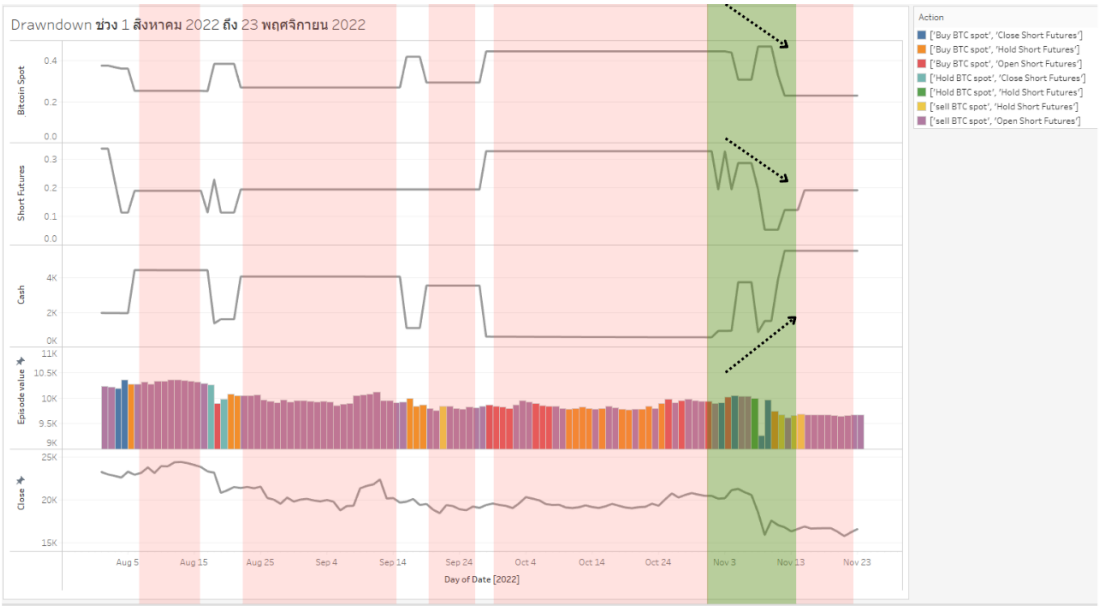
จากผลการทดสอบ มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน มีค่า Calmar Ratio เท่ากับ 1.74023 ซึ่งมีค่ามากกว่า 1 มาก แสดงถึงว่า ผลตอบแทนคุ้มค่ากับ Max Drawdown มาก

4.2 การเลือก Action ในช่วงที่เกิดต่างๆ

เพื่อเป็นการดูการเลือก Action ในช่วงที่เกิด Max Drawdown เพื่อดูการตัดสินใจของ Agent ว่าสมเหตุผลหรือไม่

4.2.1 ช่วง Max Drawdown

ซึ่งอยู่ในช่วง 1 สิงหาคม 2022 ถึง 23 พฤศจิกายน 2022 มีการเลือก Action ดังรูปที่ 4.3 ซึ่งแสดงผลโดยโปรแกรม Tableau



ภาพที่ 4.4 ภาพของการเลือก Action ในช่วง Max Drawdown

โดยมีข้อน่าสังเกตุอยู่คือ

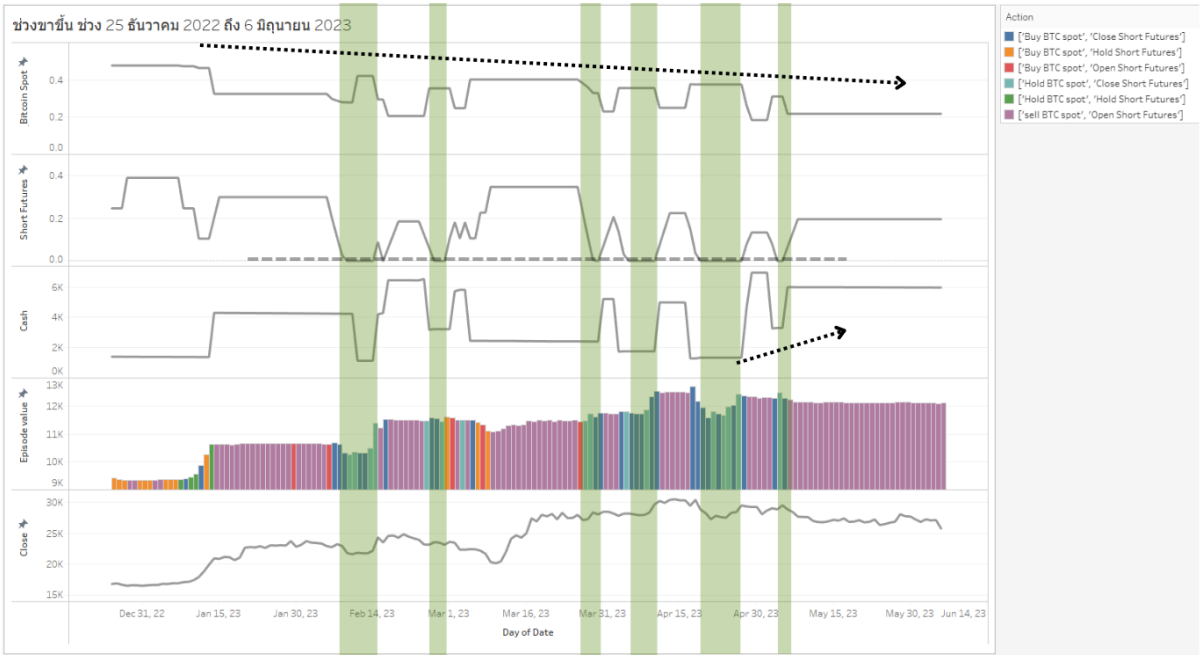
(1) ขาย BTC Spot, เปิดการขายชอร์ตสัญญาซื้อขายล่วงหน้า เป็น Action ที่ถูกเลือกเยอะสุด

(2) ช่วงเป็นเป็นสีแดง จะเป็นช่วงที่ BTC Spot,สัญญาซื้อขายล่วงหน้า ,เงินสด ไม่เปลี่ยนแปลง แม้ Agent เลือก Action เนื่องมาจาก เงื่อนไขที่ไม่สามารถขาย BTC ให้มีจำนวนต่ำกว่า การขายชอร์ตสัญญาซื้อขายล่วงหน้า ที่มีได้

(3) ช่วงสีเขียวคือช่วงที่เป็นช่วงจุดต่ำสุดของ พอร์ตการลงทุนหลังจากที่สัดส่วนนิ่งมานาน Agent เริ่มปรับสัดส่วนพอร์ตการลงทุน

4.2.2 ช่วง Uptrend

ซึ่งอยู่ในช่วง **25** ธันวาคม 2022 ถึง 6 มิถุนายน 2023 มีการเลือก Action ดังรูปที่ 4.4 ซึ่งแสดงผลโดยโปรแกรม Tableau



ภาพที่ 4.5 ภาพของการเลือก Action ในช่วง Uptrend

โดยมีข้อน่าสังเกตุอยู่คือ

(1) ขาย BTC Spot,เปิดการขายชอร์ตสัญญาซื้อขายล่วงหน้าเป็น Action ที่ถูกเลือกเยอะสุด

(2) ยังคงมีช่วงที่ Action ไปแล้วติดเงื่อนไขบางอย่างอยู่

(3) ช่วงสีเขียวคือช่วงที่ Agent เลือกที่จะไม่ถือการขายชอร์ตสัญญาซื้อขายล่วงหน้าเลย

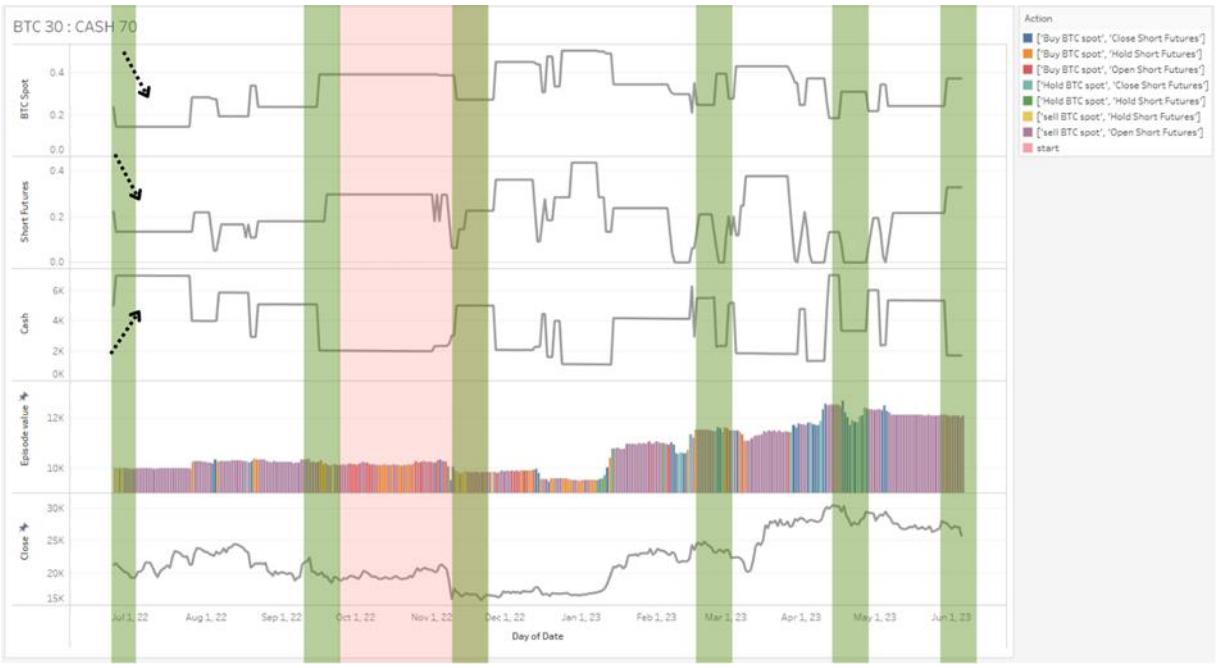
(4) Agent ถือ Bitcoin Spot เป็นแนวโน้มที่ลดลงและในช่วงหลังก็เริ่มถือเงินสดในสัดส่วนที่สูง

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

จากผลการศึกษาในบทที่ 4 ในการประยุกต์ใช้ Reinforcement learning ด้วยวิธีการ Double-Deep Q- Network มาใช้ในการปรับสมดุลย์ Bitcoin และ สัญญาซื้อขายล่วงหน้า ใน Multi-Asset Mode ซึ่งได้ Agent ที่สามารถนำไปพัฒนาต่อเป็น Trading Bot ได้แต่ถ้าหาก User ต้องการที่จะปรับอัตราส่วนเริ่มแรกที่ไม่ใช่อัตราส่วน Bitcoin 50% : เงินสด 50% Agent ยังทำงานได้มีผลเป็นอย่างไรจึงทำการทดสอบเพิ่มเติมกับ อัตราส่วน 2 แบบ คือ อัตราส่วน Bitcoin 30% : เงินสด 70% ซึ่งให้เป็นตัวแทนความเสี่ยงน้อย และ อัตราส่วน Bitcoin 70% : เงินสด 30% % ซึ่งให้เป็นตัวแทนความเสี่ยงมาก

5.1 อภิปรายผลการวิจัย

5.1.1 อัตราส่วน Bitcoin 30%: เงินสด 70%



ภาพที่ 5.1 อัตราส่วน Bitcoin 30% : เงินสด 70%

จากภาพที่ 5.1 มีข้อน่าสังเกตุดังนี้

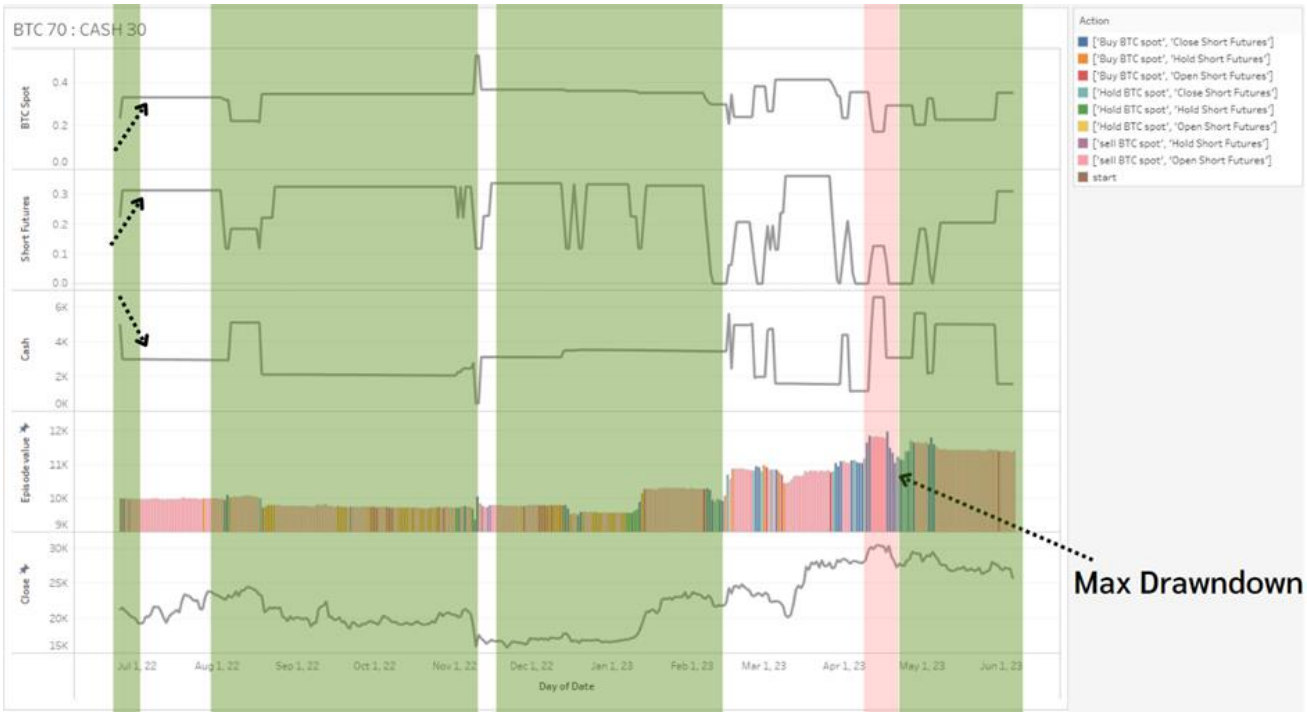
(1) มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน อยู่ที่ 12,089.80$

(2) ช่วงสีแดงคือช่วงที่เกิด Max Drawdown ซึ่งเป็นช่วงเดียวกันกับ Ratio 50%:50%

(3) ช่วงสีเขียวคือช่วงที่มีการปรับสัดส่วนที่ต่างออกไปจาก Ratio 50%:50%

(4) Action ที่เลือกทั้งหมด 7 Action ซึ่งเหมือนกับ Ratio 50%:50% Action ทั้ง 7 แบบคือ (1)'Buy BTC spot', 'Close Short Futures' (2) 'Buy BTC spot', 'Hold Short Futures' (3)'Buy BTC spot', 'Open Short Futures' (4) 'Hold BTC spot', 'Close Short Futures' (5) 'Hold BTC spot', 'Hold Short Futures' (6) 'Sell BTC spot', 'Hold Short Futures' (7) 'Sell BTC spot', 'Open Short Futures'

5.1.2 อัตราส่วน Bitcoin 70% : เงินสด 30%



ภาพที่ 5.2 อัตราส่วน Bitcoin 70% : เงินสด 30%

จากภาพที่ 5.2 มีข้อน่าสังเกตุดังนี้

(1) มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน อยู่ที่ 11,393.37 $

(2) ช่วงสีแดงคือช่วงที่เกิด Max Drawdown ซึ่งเป็นคนละช่วงกับอัตราส่วนอื่นๆ

(3) ช่วงสีเขียวคือช่วงที่มีการปรับสัดส่วน ที่ต่างออกไปจาก Ratio 50:50

(4) Agent มีการเลือก Action เพิ่มเติมที่แตกต่างออกไปจาก Ration 50%:50% และ 30%:70% คือ Hold BTC Spot , Open Short Futures Action จึงมี 8 แบบ คือ

(1)'Buy BTC spot', 'Close Short Futures'

(2) 'Buy BTC spot', 'Hold Short Futures'

(3)'Buy BTC spot', 'Open Short Futures'

(4) 'Hold BTC spot', 'Close Short Futures'

(5) 'Hold BTC spot', 'Hold Short Futures'

(6)Hold BTC Spot , Open Short Futures Action

(7) 'Sell BTC spot', 'Hold Short Futures'

(8) 'Sell BTC spot', 'Open Short Futures'

ซึ่ง Agent ยังสามารถทำการปรับสมดุลย์ ได้โดยยู่โดยอัตราส่วน Bitcoin 30% : เงินสด 70% มีมูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน ใกล้เคียงกับ อัตราส่วน Bitcoin 50% : เงินสด 50% ช่วง Drawdown อยู่ในช่วงเดียวกัน Action ทั้งหมดที่เลือกก็เหมือนกัน คือ 7 แบบ แต่สำหรับอัตราส่วน อัตราส่วน Bitcoin 70% : เงินสด 30% % มีความแตกต่างออกไป โดย มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน ค่อนข้างต่ำเมื่อเทียบกับ อัตราส่วน Bitcoin 50% : เงินสด 50% ช่วงการเกิด Drawdown ก็เกิดขึ้นคนละช่วงเวลา และ เลือก Action ถึง 8 แบบ โดยเลือก Action แบบ Hold BTC Spot , เปิดการขายชอร์ตสัญญาซื้อขายล่วงหน้า เพิ่มเข้ามา

5.1.3 การวัดผลผ่านตัววัดผลการดำเนินงานของพอร์ตการลงทุนในอัตราส่วนต่างๆ และ Buy And Hold

ตารางที่ 5.1 เปรียบเทียบการวัดผลในแต่ละอัตราส่วนต่างๆ และ Buy And Hold

Measures \ Ratio (BTC:CASH)	70:30 (เสี่ยงมาก)	50:50 (เสี่ยงกลาง)	30:70 (เสี่ยงน้อย)	Buy And Hold
Annualized Volatility	16.17547%	20.35683%	17.42254%	51.57076%
Max Drawn Down	7.62273%	10.53967%	8.80817%	35.43208%
Cumulative Returns	13.93372%	21.00965%	20.89800%	21.14406%
Annualize Sharpe Ratio	0.72110	0.90624	1.02263	0.57746
Calmar Ratio	1.47789	1.74023	2.06965	0.52144

จากตารางที่ 5.1 สามารถสรุปได้ดังนี้ ผลตอบแทนแบบอัตราส่วน 50:50 น้อยกว่า Buy and Hold แต่ได้ Drawdown ที่น้อยกว่า เมื่อเพิ่มการถือ Bitcoin เป็นอัตราส่วน 30:70 จะได้ Annualize-Sharpe Ratio และ Calmar Ratio ที่ดี เมื่อเพิ่มการถือ Bitcoin เป็นอัตราส่วน 70:30 ผลที่ออกมาคือได้ Max Drawdown และ Annualized Volatility ที่น้อยกว่าอัตราส่วนอื่น

5.2 สรุปผลการศึกษา

การใช้สัญญาซื้อขายล่วงหน้า เข้ามาป้องกันความเสี่ยงสามารถช่วยในการลดของมูลค่าสินทรัพย์สุทธิได้ โดย Agent ถือ การขายชอร์ตสัญญาซื้อขายล่วงหน้า ในช่วงขาลงเพิ่มขึ้นตามการซื้อ Bitcoin Spot เพิ่มขึ้น และ ในช่วงขาขึ้นก็มีบางช่วงที่ไม่ป้องกันความเสี่ยงเลย

สำหรับอัตราส่วนที่ BTC: CASH ที่ 50:50 และ 30:70 มีมูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน ใกล้เคียงกัน12,100.96$ และ 12,089.80$ และมี Action ที่ถูกเลือก เหมือนกัน คือ 7 Action ซึ่งเมื่อลดการถือ BTC ลงจะทำให้ Annualized Sharpe Ratio และ Calmar Ratio มีสัดส่วนสูงขึ้น

70:30 มี มูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน 11,393.37 $ และ มี Action ที่ถูกเลือกเพิ่มเข้ามา คือ ถือ BTC Spot , เปิดขายชอร์ตสัญญาซื้อขายล่วงหน้า การเพิ่ม BTC เพิ่มขึ้นจะทำให้ได้ Max Drawdown และ Volatility ลดลง เพราะมีการ ป้องกันความเสี่ยงเริ่มต้นที่ 95% ของ BTC ที่มี

5.3 ข้อเสนอแนะ

5.3.1 ใช้ Algorithm อื่นๆ แทน Double Deep Q Learning เช่น Multi-Agent PPO, A2C

5.3.2 เพิ่ม Asset ต่างๆเข้าไป เช่น ทอง ,หุ้น, Alt Coin อื่นๆ เช่น XRP,ETH

5.3.3 ปรับสัดส่วนปรับสมดุลย์ แทน k = 30% เช่น ใช้การปรับสัดส่วน k แบบอัติโนมัติตามความผันผวน

5.3.4 การกำหนดอัตราส่วนสูงสุดของการขายชอร์ตที่ไม่ได้จำกัดไว้ที่ร้อยละ 95 ของ Bitcoin spot ซึ่งสามารถเพื่อให้ Agent ป้องกันความเสี่ยงได้อย่างเหมาะสม

บรรณานุกรม

บรรณานุกรม

[1 Sutta Sornmayura (2017). Robust financial trading system with Deep Q Network (DQN). http://repository.nida.ac.th/handle/662723737/4077.

[2] Lim, Q.Y.E., Cao, Q. & Quek, C. Dynamic portfolio rebalancing through reinforcement learning. Neural Comput & Applic 34, 7125–7139 (2022). https://doi.org/10.1007/s00521-021-06853-3

[3] K. Kumlungmak and P. Vateekul, "Multi-Agent Deep Reinforcement Learning With Progressive Negative Reward for Cryptocurrency Trading," in IEEE Access, vol. 11, pp. 66440-66455, 2023, doi: 10.1109/ACCESS.2023.3289844.

[4] Hado van Hasselt , Arthur Guez , David Silver “Deep Reinforcement Learning with Double Q-Learning” https://doi.org/10.1609/aaai.v30i1.10295

[5] Pat Tong Chio ”A comparative study of the MACD base trading strategies: evidence from the US stock market” https://arxiv.org/abs/2206.12282.

[6] John F. Ehlers “ Cycle Analytics for Traders: Advanced Technical Trading Concepts ” ,ISBN: 978-1-118-72851-2, November 2013,Wiley,Page 159-164

[7] G. Kim, M. Kim, B. Kim and H. Lim, "CBITS: Crypto BERT Incorporated Trading System," in *IEEE Access*, vol. 11, pp. 6912-6921, 2023, doi:10.1109/ACCESS.2023.3236032.

[8] Andrew Barto ,Richard S. Sutton “Reinforcement Learning: an Introduction” ISBN: 978-0-262-0392-4,2020, The MIT Press,Page 48

[9] Berend Jelmer Dirk Gort, Xiao-Yang Liu, Xinghang Sun, Jiechao Gao, Shuaiyu Chen, Christina Dan Wang. Deep Reinforcement Learning for Cryptocurrency Trading: Practical Approach to Address Backtest Overfitting.https://doi.org/10.48550/arXiv.2209.05559.

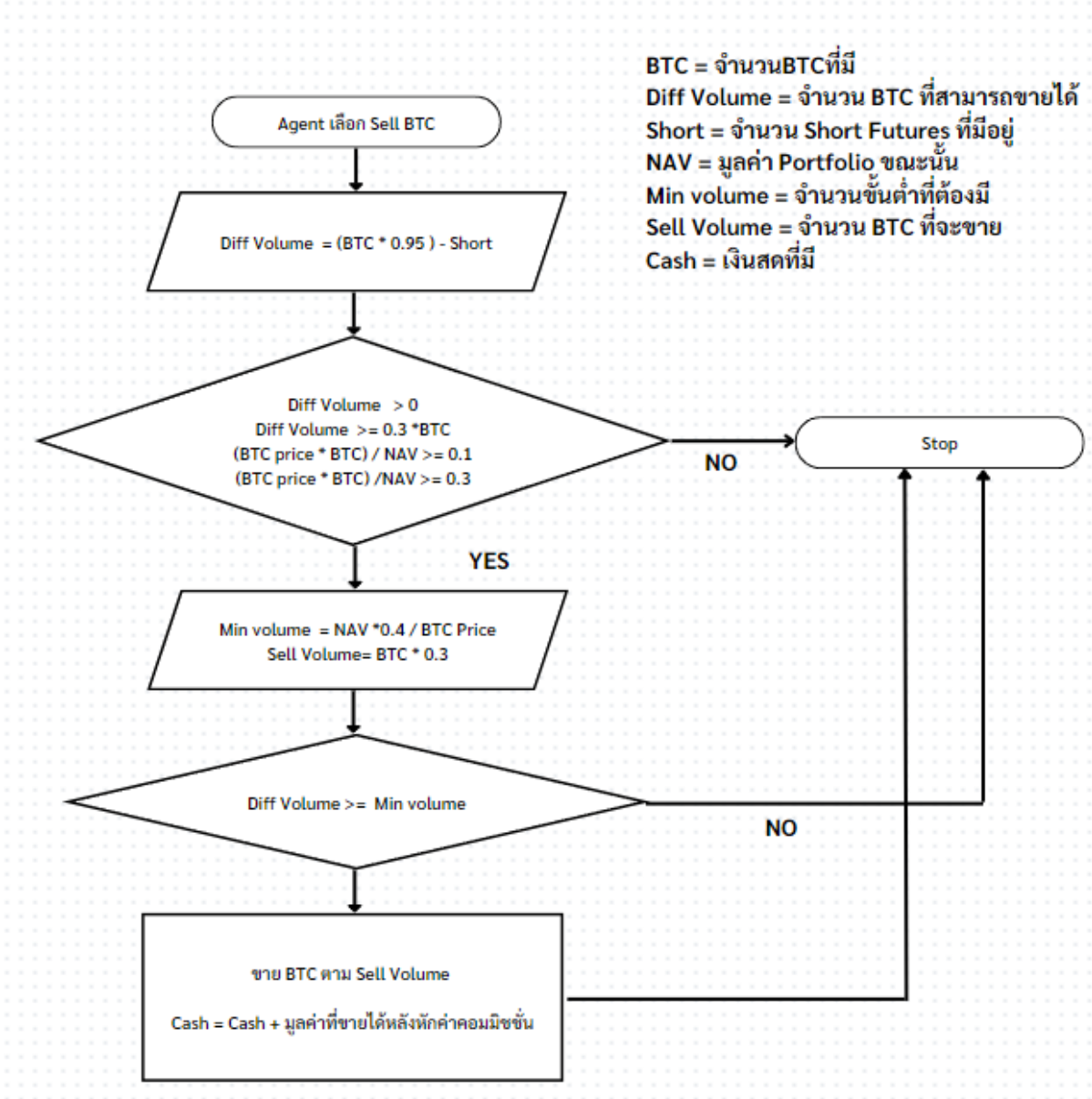
บรรณานุกรม (ต่อ)

[10] Berend Jelmer Dirk Gort, Xiao-Yang Liu, Xinghang Sun, Jiechao Gao, Shuaiyu Chen, Christina Dan Wang.Deep Reinforcement Learning for Cryptocurrency Trading: Practical Approach to Address Backtest Overfitting.https://doi.org/10.48550/arXiv.2209.05559.

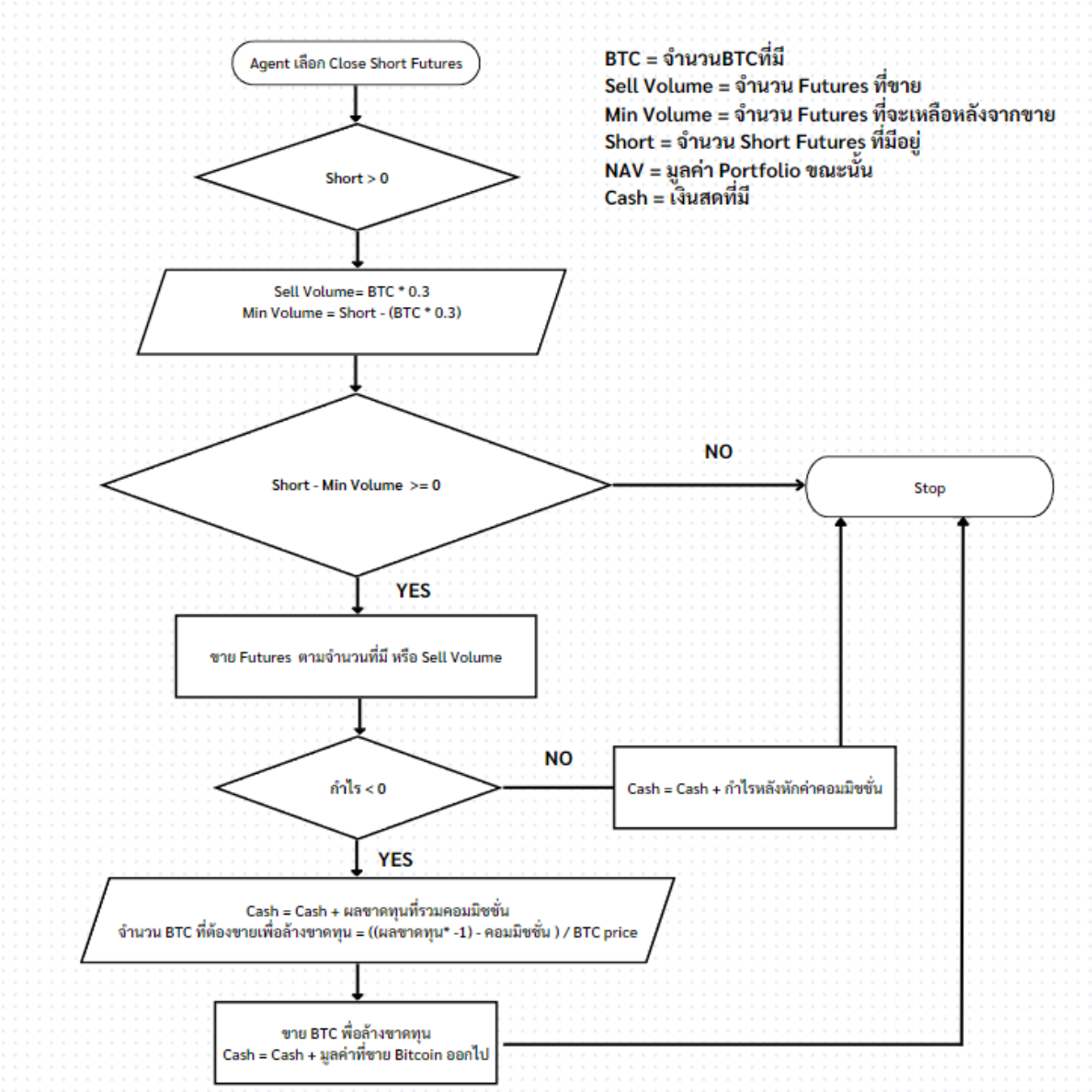
[11] Gang Huang, Xiaohua Zhou, Qingyang Song. Deep reinforcement learning for portfolio management.https://doi.org/10.48550/arXiv.2012.13773.

ภาคผนวก

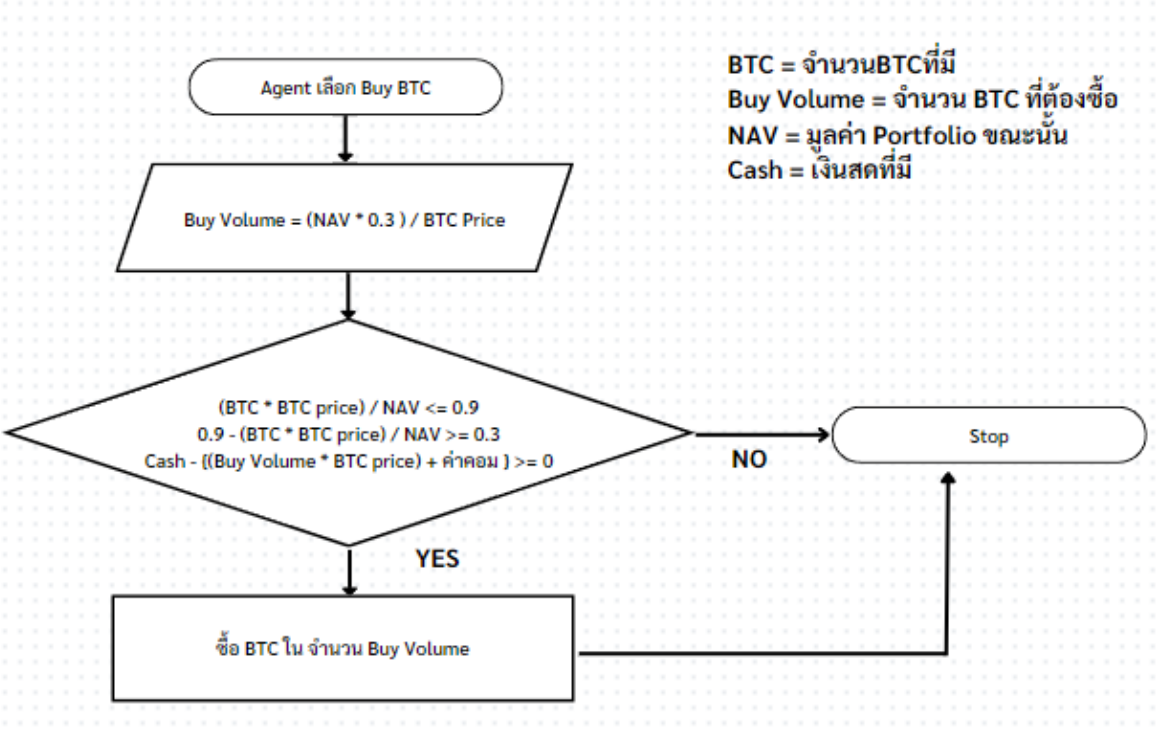
1.Flowchart เมื่อ Agent เลือกขาย Bitcoin Spot



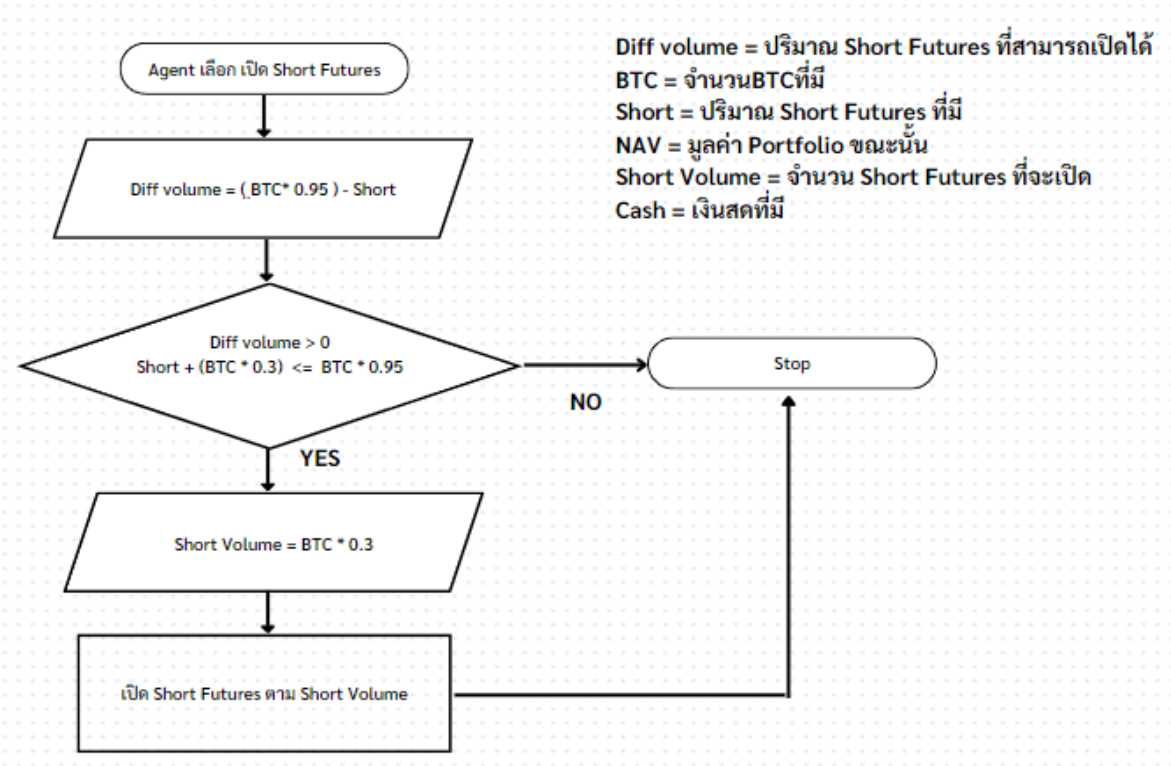
2.Flowchart เมื่อ Agent เลือกปิด การขายชอร์ตสัญญาซื้อขายล่วงหน้า



3. Flowchart เมื่อ Agent เลือกซื้อ Bitcoin Spot



4. Flowchart เมื่อ Agent เลือกเปิด การขายชอร์ตสัญญาซื้อขายล่วงหน้า



ประวัติผู้เขียน

ชื่อ-นามสกุล ธีรพันธ์ จันทร์ปราโมทย์

ประวัติการศึกษา

พ.ศ 2554 Bachelor of Economics, B. ECON. (1st class honors)
Major in Industrial Economics
มหาวิทยาลัยรามคำแหง

ประวัติการทำงาน

พ.ศ. 2565 Block Trade Officer at Beyond Securities PLC.

พ.ศ. 2564 Investment consultant at Beyond Securities PLC.

พ.ศ. 2559 Investment consultant at Maybank (TH) securities PLC.