

การรีบาลานซ์ในบิทคอยน์สปอตและป้องกันความเสี่ยงด้วยสัญญาซื้อขาย
ล่วงหน้าผ่านการเรียนรู้เชิงลึกแบบเสริมกำลัง

REBALANCING IN BITCOIN SPOT AND RISK MITIGATION WITH
FORWARD CONTRACT AGREEMENTS THROUGH DEEP
REINFORCEMENT LEARNING

ธีรพันธ์ จันทรปราโมทย์¹

ดร.ธนภัทร มังคะจิตร²

บทคัดย่อ

ในช่วงหลายปีที่ผ่านมาสกุลเงินดิจิทัล(Cryptocurrency)มีการพัฒนาไปสู่สินทรัพย์ทางการเงินที่มีการซื้อขายทั่วโลกเนื่องจากมีลงทุนที่ได้รับผลตอบแทนจำนวนมากมหาศาลเพื่อการลงทุนในสกุลเงินดิจิทัลอย่างมีประสิทธิภาพ จึงมีงานวิจัยจำนวนมากที่เสนอแนวทางในการลงทุนในสกุลเงินดิจิทัล โดยใช้เทคนิคทางด้านการเรียนรู้เชิงลึกแบบเสริมกำลัง (Reinforcement learning) เพื่อทำการสอนตัวแทน (Agent) ให้เรียนรู้พฤติกรรมของตลาดในหลากหลายรูปแบบ รวมถึงการปรับสมดุลพอร์ตการลงทุน (Rebalancing) เพื่อตอบโต้ภัยเป้าหมายการลงทุนและรักษาระดับความเสี่ยงที่เรายอมรับได้ แต่อย่างไรก็ตามมีเหตุการณ์ระบบของเหรียญสกุลเงินดิจิทัลถูกโจมตีอย่างต่อเนื่องส่งผลต่อความเชื่อมั่นและความผันผวนของค่าสกุลเงินดิจิทัล ดังนั้นงานวิจัยนี้จึงนำเสนอการแนวทางการลงทุนในสกุลเงินดิจิทัลที่มีประสิทธิภาพโดยการป้องกันความเสี่ยง (Hedging) ด้วยการชอร์ตสัญญาซื้อขายล่วงหน้า (FUTURES) เพื่อลดการขาดทุนในกรณีที่สกุลเงินดิจิทัลอ่อนตัว งานวิจัยนี้ทำการปรับสมดุลพอร์ตการลงทุนระหว่าง แบบเสริมกำลัง (Reinforcement learning) เพื่อทำการสอนตัวแทน (Agent) ให้เรียนรู้พฤติกรรมของตลาดในหลากหลายรูปแบบรวมถึงการปรับสมดุลพอร์ตการลงทุน (Rebalancing) เพื่อตอบโต้ภัยเป้าหมายการลงทุนและรักษาระดับความเสี่ยงที่เรายอมรับได้ แต่อย่างไรก็ตามมีเหตุการณ์ระบบของเหรียญสกุลเงินดิจิทัลถูกโจมตีอย่างต่อเนื่องส่งผลต่อความเชื่อมั่นและความผันผวนของค่าสกุลเงินดิจิทัล

¹ นักศึกษาหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่

² ที่ปรึกษาสารนิพนธ์หลัก

ดังนั้นงานวิจัยนี้จึงนำเสนอการแนวทางการลงทุนในสกุลเงินดิจิทัลที่มีประสิทธิภาพโดยการป้องกันความเสี่ยง (Hedging) ด้วยการชอร์ตสัญญาซื้อขายล่วงหน้า (FUTURES) เพื่อลดการขาดทุนในกรณีที่สกุลเงินดิจิทัลอ่อนตัว งานวิจัยนี้ทำการปรับสมดุลพอร์ตการลงทุนระหว่าง เงินสด (USDT/USD), สกุลเงินดิจิทัล

Bitcoin (BTC/USD) และ การขายชอร์ตสัญญาซื้อขายล่วงหน้า (Short Futures Contracts) ทั้งนี้ได้ใช้เทคนิคการเรียนรู้แบบเสริมกำลังแบบ Double Deep Q Network ร่วมกับ Multi-scale Continuous Loss เพื่อไม่ให้ตัวแทนกระทำการตัดสินใจผิดพลาด ซึ่งลดโอกาสการขาดทุนอย่างต่อเนื่องของพอร์ตการลงทุน จากผลการทดลองในช่วงเดือน มิถุนายน 2022 ถึง เดือน มิถุนายน 2023 โดยกำหนดสัดส่วนของพอร์ตการลงทุนเป็นเงินสด 50%, Bitcoin 50% และป้องกันความเสี่ยง 95% ของบิตคอยน์ที่มีอยู่พบว่าให้ผลกำไรตอบแทนใกล้เคียงกับกลยุทธ์ซื้อ Bitcoin และถือไว้ (Buy & Hold) ทั้งนี้เนื่องจากเป็นช่วงตลาดขาขึ้น

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

คำสำคัญ: การลงทุนในสกุลเงินดิจิทัล, การปรับสมดุลพอร์ตการลงทุน, การลดความเสี่ยงในการลงทุน, การชอร์ตสัญญาซื้อขายล่วงหน้า, การเรียนรู้เชิงลึกแบบเสริมกำลัง

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

บทนำ

การลงทุนแบบ Automate เริ่มเข้ามามีบทบาทในตลาดเงินมากขึ้น ทั้งในกองทุน Hedge Fund ต่างๆ รวมถึงนักลงทุนรายย่อยทั่วไป ซึ่งมีการให้บริการอยู่ในบาง Cryptocurrency Exchange การปรับสมดุล (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) ใช้ ดัชนี CVIX (Cryptocurrency Volatility Index) เข้ามาเป็นตัวควบคุมความเสี่ยงโดยหาก CVIX สูงเกินไปจากระดับที่กำหนดไว้ก็จะหยุดการซื้อขาย แล้วขาย Cryptocurrency ที่มีอยู่ทั้งหมดแล้วจะกลับมาทำการซื้อขายต่อเมื่อ CVIX กลับมาอยู่ในระดับปกติ

Giorgio Lucarelli, Matteo Borrotti (2020) มีการใช้การตัดขาดทุนเมื่อขาดทุนมากกว่า 2.3% และ ทำกำไรเมื่อกำไรมากกว่า 6.5% ในการซื้อขาย Cryptocurrency

Gang Huang, Xiaohua Zhou and Qingyang Song (2020) ประยุกต์ใช้ Deep Reinforcement Learning ในการซื้อขายหุ้นในดัชนี CSI300 โดยให้ Agent สามารถเลือกขายชอร์ตหุ้น (Short Sell) เพื่อทำกำไรหากหุ้นถูกประเมินว่ามีมูลค่าสูงเกินมูลค่าที่แท้จริง ซึ่ง Agent สามารถเลือกกระทำได้เพียงอย่างเดียวหนึ่งกับ

หุ้นว่าจะซื้อ หรือ ขายชอร์ต แต่เนื่องจากตลาด Cryptocurrency เป็นตลาดที่มีความผันผวนมาก การใช้ดัชนี หรือ ตั้งจุดตัดขาดทุนอาจเกิดการขาดทุนโดยไม่จำเป็นหากเป็นระดับที่ไม่เหมาะสม หรือประยุกต์ใช้ขายชอร์ต กับตลาด Cryptocurrency เพื่อทำกำไรในช่วงที่ราคาลง ก็มีโอกาสดูทุนในพอร์ตการลงทุนหากมีสัดส่วนมาก เกินไปในช่วงที่ราคาขึ้น

ดังนั้น ในการทำการปรับสมดุล ผู้เขียนจึงเพิ่มการขายชอร์ตสัญญาซื้อขายล่วงหน้า แบบไม่มีวันหมดอายุเข้าไปในพอร์ตการลงทุน ที่ทำการปรับสมดุลเหรียญ Bitcoin เพื่อป้องกันความเสี่ยงที่จะเกิดขึ้นกับ พอร์ตการลงทุน โดยอัลกอริทึมที่นำมาใช้คือ Double Deep Q Network ที่ใช้การคิด Reward แบบ Multi-scale continuous loss (MSCL) เพื่อลดโอกาสการลดลงอย่างต่อเนื่องของมูลค่าสินทรัพย์สุทธิซึ่งมีแนวคิดมาจากการ เพิ่มบทลงโทษที่มากขึ้นในแต่ละครั้งถ้าทำผิดแบบเดิมซ้ำโดยให้เริ่มต้นพอร์ตการลงทุนที่สัดส่วน Bitcoin spot 50%:เงินสด50%ผลลัพธ์ที่ได้คือ Max Drawdown ลดลงมากเมื่อเทียบกับการซื้อแล้วถือ (Buy and 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

วัตถุประสงค์ของงานวิจัย

1. เพื่อนำเสนอการจัด พอร์ตการลงทุน ในการลงทุน Bitcoin ที่เพิ่ม การขายชอร์ตสัญญาซื้อขายล่วงหน้า เข้ามาเพื่อป้องกันความเสี่ยง
2. เพื่อนำเสนอการปรับสมดุล ด้วย Double Deep Q Network ที่ใช้การคิด Reward แบบ Multi-scale continuous- loss (MSCL) เพื่อป้องกันไม่ให้ Agent ทำผิดซ้ำแบบเดิม

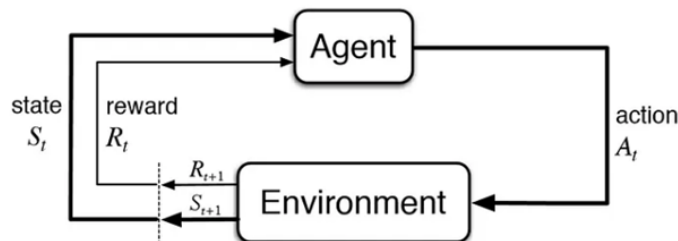
ขอบเขตงานวิจัย

1. พัฒนา Trading bot ของ Bitcoin
2. การขายชอร์ตสัญญาซื้อขายล่วงหน้า และ Spot ที่ใช้จะเป็น BTC/USDT
3. การปรับสมดุล เป็นแบบ Daily
4. ใช้ Multi Asset Mode ซึ่งทาง Binance เปิดให้สามารถนำเอาทรัพย์สินในพอร์ตการลงทุน ไปวาง เป็นหลักประกันเพื่อเปิดการขายชอร์ตสัญญาซื้อขายล่วงหน้าได้

5. สัญญาซื้อขายล่วงหน้าเป็นแบบไม่มีวันหมดอายุ (Perpetual)

ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

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



ภาพที่ 1 รูปแบบการทำงานของ Reinforcement Learning

2. Double Deep Q Network คำนวณค่า Target ด้วยการใช้ค่า Max ซึ่งคือการเลือกค่าที่มากที่สุด แล้วการประมาณค่า Q-value และ เลือก Action ของ Target ก็จะทำพร้อมกันบน Network เดียวกัน ซึ่งทำให้เกิดปัญหา 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

$$y = r + \gamma Q(s', \arg\max_a Q(s', a'; \theta); \theta' | s, a)$$

เมื่อ y คือ ค่าที่ใช้เป็น Target ใน Loss Function

r คือ Reward ของ state s และ action a

γ คือ discount rate

$Q(s', a'; \theta)$ คือ ค่า Q-value ของ state s' และ action a' ซึ่งเป็น State ถัดไป

Sutta Sornmayura (2017). Robust Forex Trading With Deep Q Network งานชิ้นนี้เป็นการศึกษาเกี่ยวกับการใช้ปัญญาประดิษฐ์ในการพัฒนาระบบการเทรด Forex ในสกุลเงิน EUR/USD และ USD/JPY โดยใช้ Deep Q Network (DQN) โดยมีการใช้ Sine wave Indicators เป็น 1 ใน State โดยผลลัพธ์จากการเทรดของ Agent ถูกนำไปเทียบกับกลยุทธ์ Buy and hold และ ข้อมูลการเทรดของเทรดเดอร์ที่เชี่ยวชาญโดยผลลัพธ์ที่ได้ Agent ได้ผลตอบแทนมากกว่ากลยุทธ์ Buy and hold และ ข้อมูลการเทรดของเทรดเดอร์ อย่างมีนัยยะสำคัญ

Lim, Q.Y.E., Cao, Q. & Quek (2021). Dynamic portfolio rebalancing through reinforcement - learning งานชิ้นนี้ การศึกษาเกี่ยวกับการทดสอบการ ปรับสมดุล 4 แบบ กับ สินทรัพย์ 3 ประเภท ด้วย Deep Q Network (DQN) โดยใน 1 สินทรัพย์จะประกอบด้วย สินทรัพย์เสี่ยงสูง,เสี่ยงกลาง,เสี่ยงน้อย สินทรัพย์ 3 ประเภท ได้แก่ (1) ดัชนีตลาดหลักทรัพย์ (2) หุ้นในตลาด S&P 500 (3) หุ้นในตลาด Nasdaq โดย Agent จะ Action โดย ปรับสมดุลพอร์ตการลงทุนเมื่อมีการทำนายว่าเกิดการกลับตัวของแนวโน้ม โดย Action มี 4 แบบ ซึ่ง Composition rate คือ

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

เมื่อ BCR คือ Base Composition Rate ซึ่งมีค่าเท่ากับ 0.1

n คือ เท่ากับจำนวนสินทรัพย์

สัดส่วนจะเปลี่ยนไปมาระหว่าง 2 ค่า คือ 0.1 และ 0.8 ซึ่งพบว่าเป็นการเปลี่ยนแปลงที่มากเกินไปจึงมีการใช้ การทยอยปรับสมดุลพอร์ตการลงทุน(Gradual Portfolio Rebalancing) ครั้งละ k% โดยค่า k ค่าที่เหมาะสมคือ 0.3 และ ใช้มีการนำไปใช้กับแบบที่ ไม่ใช่ โมเดลพยากรณ์กับ แบบใช้โมเดลพยากรณ์ ผลลัพธ์ที่ได้จากการ ปรับสมดุล ด้วยวิธีการทยอยปรับสมดุลพอร์ตการลงทุนแบบมีโมเดลพยากรณ์ได้ ผลตอบแทนมากที่สุด และ Max Drop น้อยที่สุด

Kittiwinn Kumlungmak, Peerapon Vateekul (2023) Multi-Agent Deep Reinforcement Learning With Progressive Negative Reward for Cryptocurrency Trading ได้นำเสนอเกี่ยวกับการใช้ Multi-Agent Proximal Policy Optimization ใน Cryptocurrency 6 ชนิด ADAUSDT, BNBUSDT, BTCUSDT, ETHUSDT, XRPUSDT, 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

สำหรับ Reward จะประกอบไปด้วย 2 ส่วนคือ Local Reward และ Global Reward โดย Local Reward จะใช้กับการคำนวณ Reward เฉพาะใน เหรียญนั้นๆ Global Reward จะใช้กับการคำนวณ Agent ใน ทุกๆตัว โดยงานวิจัยชิ้นนี้ได้เน้นการนำเสนอ Multi-Scale Continuous Loss Reward (MSCL) ซึ่งคือ

$$r_MSCL(S_t, A_t, S_(t+1)) = \mu \cdot |n_loss - n_win| \cdot r_(\text{portfolio}, t)$$

โดยให้ $\mu \geq 0$ โดยมีจุดประสงค์เพื่อให้ Agent ลดการทำผิดพลาดในสิ่งเดิม ๆ ผ่านการเพิ่มค่าติดลบใน Reward โดยที่เงื่อนไข คือ

- (1) มูลค่า มูลค่าสินทรัพย์สุทธิ ขณะนั้น น้อยกว่า มูลค่าสินทรัพย์สุทธิตอนเริ่มต้น
- (2) จำนวนครั้งของการเทรดขาดทุนมากกว่ากำไร
- (3) รางวัลจากการเปลี่ยนแปลงมูลค่าของพอร์ตการลงทุน ≤ 0

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

ระเบียบวิธีวิจัย

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

1. การจำลอง Environment ของ Cryptocurrency Exchange เพื่อให้ Agent สามารถเรียนรู้เพื่อเทรดใน สินค้า Bitcoin และ การขายชอร์ตสัญญาซื้อขายล่วงหน้า ได้ Environment ต้องมีส่วนประกอบดังนี้

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 (1) Market Data ราคาปิด Bitcoin spot , ราคาปิด สัญญาซื้อขายล่วงหน้าของ Bitcoin , MACD , MACD signal , Even better Sinewave , ผลต่างราคาปิด Bitcoin spot กับ สัญญาซื้อขายล่วงหน้าของ Bitcoin (2) Portfolio Data Portfolio Data คือ สัดส่วนมูลค่าของ Bitcoin spot ต่อมูลค่า พอร์ตการลงทุน , สัดส่วนเงินสด ต่อมูลค่า พอร์ตการลงทุน, สัดส่วนสัญญาซื้อขายล่วงหน้าของ Bitcoin ต่อ จำนวน Bitcoin spot

1.2 Action มี 3 แบบ คือ Buy , Hold , Sell ใน 1 Asset ซึ่งจะกระทำกับ Bitcoin spot และ การขายชอร์ต สัญญาซื้อขายล่วงหน้าของ Bitcoin พร้อมกัน ดังนั้น Action ที่เกิดขึ้นได้ ทั้งหมด แบบ คือ



ภาพที่ 2 แสดง Action ทั้งหมด 9 รูปแบบ

1.3 Reward Agent จะเรียนรู้ผ่าน Reward จากการ Action ในแต่ละ State โดย Reward ที่ใช้ในงาน สารนิพนธ์นี้ ประกอบด้วย 2 ส่วน

(1) NAV reward คือ มูลค่าสินทรัพย์ใน พอร์ตการลงทุน (Net Asset Value ; NAV) ณ ช่วงเวลา นั้นเทียบกับช่วงก่อน โดยคำนวณจาก

$$\text{NAV reward} = (\text{NAV}_{(t+1)}) / (\text{NAV}_t) - 1$$

เมื่อ $\text{NAV}_{(t+1)}$ คือ มูลค่าสินทรัพย์สุทธิ ของ State ถัดไปหลังจากมีการ Action ไปแล้ว

NAV_t คือ มูลค่าสินทรัพย์สุทธิ ของ State ก่อนมีการ Action

(2) Multi-scale continuous loss (MSCL) Reward เพื่อลดโอกาสการลดลงอย่างต่อเนื่องของมูลค่า สินทรัพย์สุทธิ ซึ่งมีแนวคิดมาจากการเพิ่มบทลงโทษที่มากขึ้นในแต่ละครั้ง ถ้าทำผิดแบบเดิมๆ ซ้ำๆ ดังนั้นหาก เกิดเงื่อนไขเหล่านี้ขึ้นพร้อมกัน Reward จะติดลบมากขึ้น โดยคำนวณจากสมการด้านล่าง

$$\text{MSCL reward} = \text{NAV reward}_t + [((\text{NAV}_t) / (\text{NAV}_{\text{start}}) - 1) \cdot (N_{\text{win}} - N_{\text{loss}}) \cdot (\text{NAV reward}_t)]$$

เมื่อ N_{win} คือ จำนวนครั้งที่ชนะ

N_{loss} คือ จำนวนครั้งที่แพ้

โดย MSCL Reward จะถูกใช้คำนวณเมื่อเข้าเงื่อนไขทั้ง 3 ข้อดังนี้

- (1) มูลค่า มูลค่าสินทรัพย์สุทธิ ณ Timestep นั้น น้อยกว่า มูลค่าสินทรัพย์สุทธิ ที่ Initial
- (2) จำนวนครั้งการเทรดขาดทุนมากกว่ากำไร
- (3) มูลค่าสินทรัพย์สุทธิ Reward ณ Timestep นั้น น้อยกว่าเท่ากับ 0

1.4 ขีดจำกัดหรือเงื่อนไขของการทดลอง

(1) ใช้ base composition rates (BCR) และ composition rates (CR) ในการควบคุมสัดส่วนขั้นต่ำและสัดส่วนสูง (ขีดจำกัด) สูงของ Bitcoin spot กับ เงินสด เท่านั้น เพื่อป้องกันไม่ให้มี Bitcoin หรือ เงินสด 100% BCR มีค่าเท่ากับ 0.1 และ CR จึงมีค่าเท่ากับ 0.9

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

(3) ใช้ Multi Asset Mode ซึ่งทาง Binance เปิดให้สามารถนำเอาทรัพย์สินใน พอร์ตการลงทุนไปวางเป็นหลักประกันเพื่อเปิดสัญญา ขายขอร์ตสัญญาซื้อขายล่วงหน้า เพื่อ ป้องกันความเสี่ยง ได้โดยคิดมูลค่าเท่ากับ 95% ของมูลค่าทรัพย์สินที่นำมาวาง ดังนั้น ปริมาณ การขายขอร์ตสัญญาซื้อขายล่วงหน้า ที่มีสูงสุด ณ Time step หนึ่งจะเท่ากับ $\text{Max Short Futures Position} = 0.95 \cdot \text{Bitcoin spot}$

(4) เมื่อมีการปิดสัญญา การขายขอร์ตสัญญาซื้อขายล่วงหน้า แล้วขาดทุนก็จะทำการขาย Bitcoin spot ชดเชยผลขาดทุนนั้นเพื่อให้เงินสดติดลบ

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

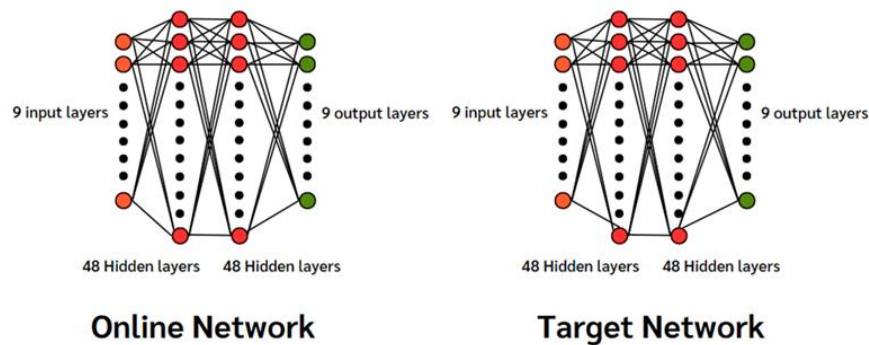
(6) ค่าคอมการขายขอร์ตสัญญาซื้อขายล่วงหน้า ใช้ Taker Rate 0.04% ส่วน Spot ใช้ 0.10%

(7) การถือ การขายขอร์ตสัญญาซื้อขายล่วงหน้า จะถูกคิด Funding Rate เนื่องจากสัญญา การขายขอร์ตสัญญาซื้อขายล่วงหน้า ที่ใช้เป็น สัญญาซื้อขายล่วงหน้า แบบที่ไม่หมดอายุ (Perpetual Futures)ซึ่งราคา สัญญาซื้อขายล่วงหน้า ในตลาดสามารถมากกว่า หรือ น้อยกว่า Bitcoin Spot ได้ Funding Rate จึงเป็นตัวปรับให้ราคา สัญญาซื้อขายล่วงหน้า อยู่บนพื้นฐานราคาจริง

(8) ระบบนับกำไรขาดทุนแบบ FIFO

1.5 Algorithm และ โครงสร้างของ Network ใช้ Algorithm ที่ชื่อว่า Double Deep Q Network ซึ่งช่วยลดปัญหา Overestimation ใน Deep Q Network โดยมีโครงสร้าง ประกอบด้วย

- (1) Online Network และ (2) Target Network



ภาพที่ 3 ภาพแสดงโครงสร้างของ Network ใน Double Deep Q Network

ซึ่งจากรูปที่ 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 โดยมีการใช้

(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

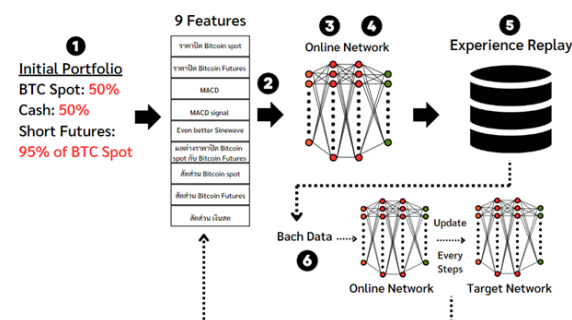
$$\theta' = \tau\theta + (1-\tau)\theta'$$

เมื่อ θ คือ ค่า Weight ของ Online Network

θ' คือ ค่า Weight ของ Target Network

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

3 การ Training ผ่าน Google Colab โดยตั้งค่า Hyper-parameters ต่างๆ ดังนี้



ภาพที่ 4 ภาพแสดงขั้นตอนการแสดงผลการ Train ของ Double Deep Q Network

(1) Alpha = 0.00001 (2) Gamma = 0.95 (3) Epsilon Decay = 0.999

(4) Batch Size = 32 (5) Initial Money = 10,000 \$

ขั้นตอนการเทรนจากภาพที่4 คือ

(1) จะเริ่มต้นด้วยการ Initial พอร์ตการลงทุน ขึ้นมาด้วยอัตราส่วน Bitcoin 50%, เงินสด 50% และ การขายชอร์ตสัญญาซื้อขายล่วงหน้า 95% ของ Bitcoin ให้ Epsilon = 1.00

(2) จากนั้นก็จะเปลี่ยนเป็น State ถัดไปโดยนำเอา State 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 มาตาม ขนาด Batch size เพื่อ Train Neural Network และ Update Target แล้วลดค่า Epsilon ด้วย การคูณกับ Epsilon Decay จากนั้นก็จะเริ่มใหม่ในข้อ 2 โดยได้ทำการ Train ไป 1800 Episodes

ผลการศึกษา

ตารางที่ 1 เปรียบเทียบการวัดผลในอัตราส่วน 50%:50% และ Buy And Hold

Ratio (BTC:CASH)	50:50	Buy And Hold
Measures		
Annualized Volatility	20.35683%	51.57076%
Max Drawn Down	10.53967%	35.43208%
Cumulative Returns	21.00965%	21.14406%
Annualize Sharpe Ratio	0.90624	0.57746
Calmar Ratio	1.74023	0.52144

นำผลที่ได้มาวัดผลกับ Portfolio Performance metrics 5 ชนิด ผลคือ Annualized Volatility และ Max Drawn Down น้อยกว่าการ Buy And Hold สะท้อนถึงความเสี่ยงที่ Agent สามารถทำได้น้อยกว่าแม้ว่า Cumulative Returns Agent ทำได้น้อยกว่า Buy And Hold แต่ก็ต่างกันเพียง 0.14% สำหรับ Annualize Sharpe Ratio ค่าที่ออกมาอาจจะไม่คุ้มค่ามากนักกับความเสี่ยง เพราะมีค่าน้อยกว่า 1 แต่ Calmar Ratio ได้ค่าออกมาถึง 1.74 เท่า ซึ่งคุ้มค่ามากเมื่อเทียบกับ Max Drawn Down ของ พอร์ตการลงทุน

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

1. การวัดผลผ่าน Portfolio Performance Metrics ในอัตราส่วนต่างๆ และ Buy And Hold

ตารางที่ 2 เปรียบเทียบการวัดผลในแต่ละอัตราส่วนต่างๆ และ Buy And Hold

Ratio (BTC:CASH) Measures	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 2 สามารถสรุปได้ดังนี้ ผลตอบแทนแบบอัตราส่วน 50%:50% น้อยกว่า ซื้อ Bitcoin และถือไว้ 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

สรุปผลการศึกษา

การใช้ สัญญาซื้อขายล่วงหน้า เข้ามา ป้องกันความเสี่ยง สามารถช่วยในการลด มูลค่าสินทรัพย์สุทธิได้ โดย Agent ถือ การขายชอร์ตสัญญาซื้อขายล่วงหน้า ในช่วงขาลงเพิ่มขึ้นตามการซื้อ Bitcoin Spot เพิ่มขึ้น และ ในช่วงขาขึ้นก็มีบางช่วงที่ไม่ ป้องกันความเสี่ยง

สำหรับอัตราส่วนที่ BTC: CASH ที่ 50%:50% และ 30%:70% มีมูลค่าสินทรัพย์สุทธิของ พอร์ตการลงทุน ใกล้เคียงกัน 12,100.96\$ 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

ข้อเสนอแนะ

1. ใช้ Algorithm อื่นๆ แทน Double Deep Q Learning เช่น Multi-Agent PPO , A2C
2. เพิ่ม Asset ต่างๆเข้าไป เช่น ทอง ,หุ้น, Alt Coin อื่นๆ เช่น XRP,ETH
3. ปรับสัดส่วน ปรับสมคูลย์ แทน $k = 30\%$ เช่น ใช้การปรับสัดส่วน k แบบอัตโนมัติตามความผันผวน
4. การกำหนดอัตราส่วนสูงสุดของ การขายชอร์ตที่ไม่ได้จำกัดไว้ที่ร้อยละ 95 ของ Bitcoin spot ซึ่งสามารถ เพื่อให้ Agent ป้องกันความเสี่ยง ได้อย่างเหมาะสม

บรรณานุกรม

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>.

- 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>.
- Gang Huang, Xiaohua Zhou, Qingyang Song. Deep reinforcement learning for portfolio management. <https://doi.org/10.48550/arXiv.2012.13773>.
- Hado van Hasselt, Arthur Guez, David Silver "Deep Reinforcement Learning with Double Q-Learning" <https://doi.org/10.1609/aaai.v30i1.10295>
- 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
- 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>
- Richard S. Sutton and Andrew G. Barto (2020). Reinforcement Learning: An Introduction second edition. The MIT Press.
- Sutta Sornmayura (2017). Robust financial trading system with Deep Q Network (DQN). <http://repository.nida.ac.th/handle/662723737/4077>.