Transferable Reinforcement Learning in Forex Trading

EEMCS, Delft University of Technology
Author: Yavuz Hancer
Supervisors: Neil Yorke-Smith, Antonis Papapantoleon, Amin Kolarijani

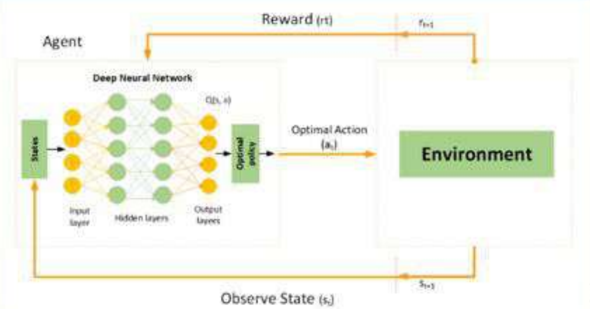
Cross-Currency Adaptation Techniques for EUR/USD and GBP/USD

1. Introduction

- **Forex Market Overview:**
  - Largest financial market globally, with daily trading volumes exceeding $6 trillion[1].
  - Major currency pairs like EUR/USD and GBP/USD are characterized by high noise, nonstationarity, and frequent regime shifts.
- **Automated Trading:**
  - Traditional supervised models have a hard time adapting online to sudden market changes.
  - Deep reinforcement learning (RL) frameworks treat trading as a Markov Decision Process (MDP)[2]:
- **Transfer Learning:**
  - Seeks to leverage knowledge (network weights or representations) from a "source" task to accelerate learning in a "target" task.
- **Research Objective**
  - Primary Research Question:

How effectively can transfer learning techniques reduce training time and improve the performance of RL agents when applied to new currency pairs?

  - Sub-Questions:
- *How do these strategies compare in terms of final trading performance, as measured by cumulative reward and Sharpe ratio, relative to training an agent from scratch?*
- *What are the trade-offs between run-time efficiency and policy effectiveness when deploying transfer learning techniques in forex trading environments?*
- **Performance Metrics:**
- *Sharpe Ratio: Measures the average excess return per unit of volatility, indicating how well the agent balances profit against risk[4].*
- *Cumulative Reward: The total sum of profits and losses accumulated by the agent over the evaluation period.*
- *Training Time: The wall-clock time required to complete the specified number of training steps.*



2. Methodology

- **Data Collection and Preprocessing:**
  - 15-minute OHLCV data for EUR/USD and GBP/USD are sourced from Dukascopy.
  - Timestamps are aligned, and any missing bars are forward-filled to maintain continuity.
  - Each pair is split chronologically into 70% training and 30% held-out evaluation to prevent look-ahead bias.
- **Feature Engineering:**
  - Market Indicators: ATR, MACD, and RSI capture volatility and momentum, providing valuable insights into market trends[5][6].
  - Temporal Patterns: Sine–cosine encodings of intra-day and intra-week cycles to reflect trading hours and weekday effects.
  - Agent Stete: The Fraction of capital currently deployed and the cumulative trade duration inform risk exposure.
- **RL Environment and Agent:**
  - Modeled as an MDP stepping at each 15-minute bar.
    - State (S): Concatenation of market and agent-state features.
    - Action (A): Discrete {-1 = Short, 0 = Hold, +1 = Long}.
    - Reward (R): Instantaneous P/L per bar.
  - DQN Architecture: A multi-layer perceptron approximates the action-value function Q(s,a), mapping state features to Q-values for each discrete action[3]. Training is stabilized via experience replay and a periodically updated target network, with actions selected using an ε-greedy policy.
- Transfer Techniques:
  - Zero-Shot Transfer: Directly evaluate the EUR/USD–trained DQN policy on GBP/USD without any further training.
  - Full Fine-Tuning: Initialize all network weights from the pretrained EUR/USD model and continue training on GBP/USD.
  - Partial Fine-Tuning: Freeze the first hidden layer (as a generic feature extractor) and fine-tune only the deeper decision layers on GBP/USD.
  - Reward Function Transfer: Retain the EUR/USD Q-network architecture and initial weights but retrain under a Sharpe ratio adjusted reward on GBP\USD.

Figure 4: Deep Q-Network (DQN) agent interacting with the environment visualized.

3. Results

Strategy	Sharpe Ratio	Cumulative Reward	Training Time (s)
Zero-Shot Transfer	-0.0061	-233.3	8.2
Full Fine-Tuning	-0.0151	-13.0	36.5
Partial Fine-Tuning	0.0057	214.5	32.8
Reward-Function Transfer	-0.0162	-1.5	38.3
From-Scratch Baseline	0.0268	749.9	37.3

Table 1: Performance comparison between different transfer techniques on GBP/USD.

- **Sample Efficiency**: Zero-shot is fastest but catastrophically underperforms, generating large losses
- **Adaptation Trade-off**: Partial fine-tuning accelerates early learning with a positive cumulative reward but still below the from-scratch baseline.
- **Overall Performance:** The from-scratch agent achieves the highest cumulative profit and Sharpe ratio[], indicating that GBP/USD requires domain-specific learning.
- **Catastrophic Forgetting:** Full fine-tuning and reward-function transfer both yield negative returns, suggesting wholesale weight updates or reward reshaping alone cannot overcome domain mismatch.

4. Conclusion

- **Limited Out-of-the-Box Transfer:** Direct zero-shot policies fail to generalize across even closely related FX pairs.
- **Partial Fine-Tune Promises:** Freezing low-level layers preserves useful market features and speeds early learning, but cannot match full retraining.
- **From-Scratch Superiority:** When sufficient data and computing resources are available, training from scratch remains the most reliable approach for performance.

5. Future Work

- Systematically compare the learning curves of each transfer method (cumulative reward vs. training steps) to quantify sample-efficiency and convergence rates.
- Run each experiment across multiple random seeds and market regimes to assess variance and ensure findings generalize beyond a single train/test split.
- Explore progressive or curriculum-based transfer, where models are adapted through a sequence of intermediate currency pairs or market regimes rather than a direct one-shot jump.

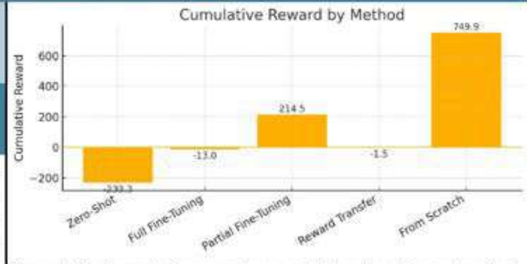


Figure 1: Final cumulative reward across training steps for each method, as reported in Table 1.

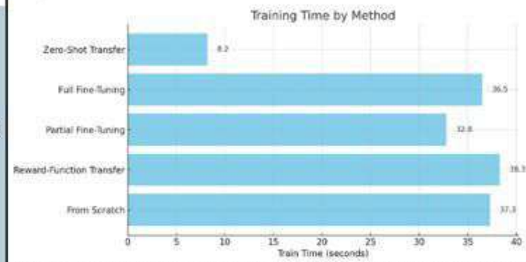


Figure 2: Final training runtime across training steps for each method, as reported in Table 1.

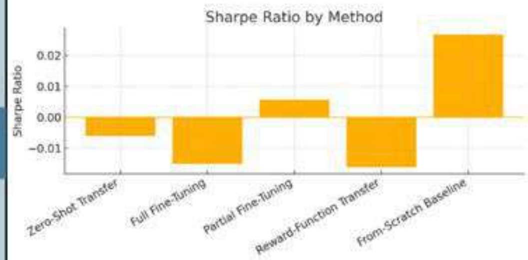


Figure 3: Final Sharpe ratio after training steps for each method, as reported in Table 1.

Resources:

1. Bank for International Settlements. Examination of foreign exchange and otc derivatives markets. BIS Triennial Central Bank Survey, 2019. https://www.bis.org.
2. Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. MIT Press, 2018.
3. Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, et al. Human-level control through deep reinforcement learning. Nature, 518(7540):529–533, 2015.
4. William F. Sharpe. Mutual fund performance. The Journal of Business, 39(1):119–138, 1966.
5. Cheol-Ho Park and Scott H. Irwin. What do we know about the profitability of technical analysis? Journal of Economic Surveys, 21(4):786–826, 2007.
6. J. Welles Wilder. New Concepts in Technical Trading Systems. Trend Research, 1978.