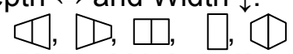
The Impact of Function Approximation Methods on Model Performance

The use of Reinforcement Learning in Algorithmic Trading

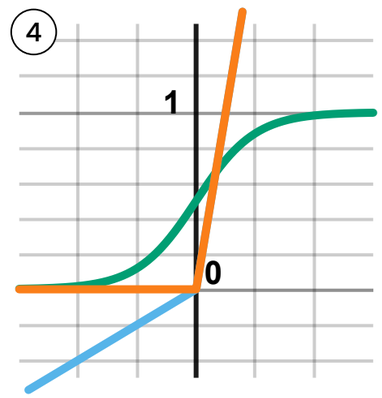
1. Background
- Forex Trading
- Reinforcement Learning
- Function Approximation

2. Research Questions
- Network Architecture (Size, Shape, Division)
- Activation Functions
- TA Features vs CNN
- L2 Regularization

3. ForexEnv
a. Observations of engineered Features.
b. Continuous actions in [-1, 1]
c. Execute trade: incl Commission and Spread, Zero Market Impact.
d. Potential Based Rewards using DP computation of V*(s).
e. Repeat.

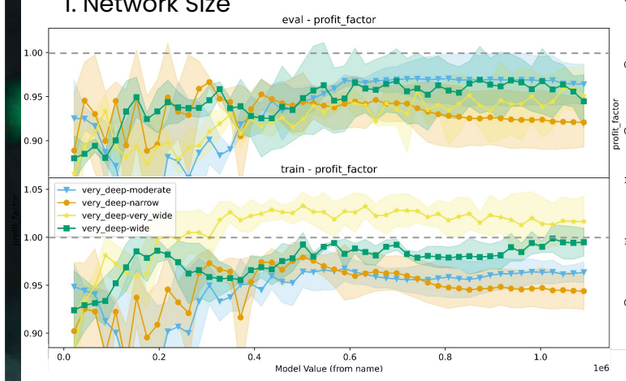
4. Experiments
Baseline: SAC, Off-policy, Actor-Critic.
1. Network Size: Depth ↔ and Width ↕.
2. Network Shapes: 
3. Network Division: 50/50, 60/40, 75/25
4. Activation Functions: (Leaky)ReLU, Sigmoid, etc
5. Technical Analysis vs Convolutional NN.
6. Weight Decay (L2 Regularization)

Train with 5 different seeds, evaluate every episode.
Data: 5 Years of 1 hour data from Dukascopy.

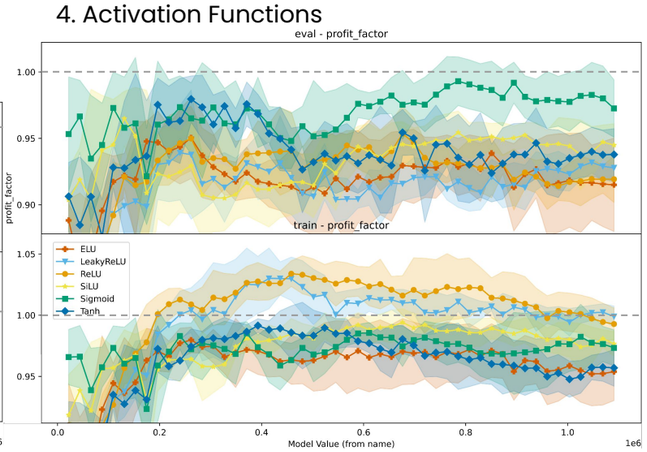


5. Results

1. Network Size



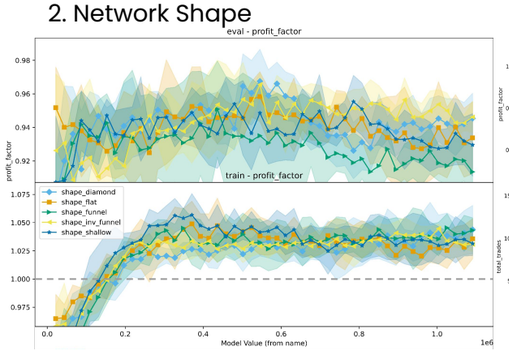
4. Activation Functions



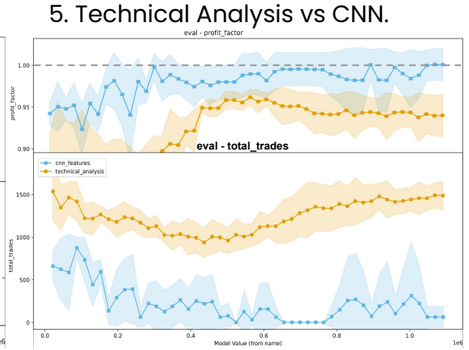
3. Network Division
Critic underfitting
Similar eval profit.

6. Weight Decay
Underfitting

2. Network Shape



5. Technical Analysis vs CNN.



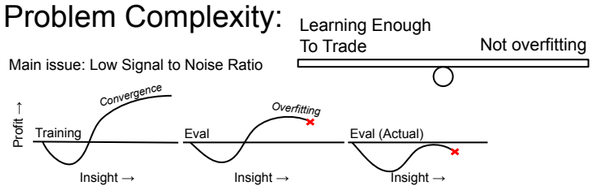
Best Performers: Sigmoid, Moderate-width, Deep Networks, TA-Features.

6. Discussion
Results interpretation
- Network Size: Balance
- Network Size / Division: Fixed parameters
- Activation Functions: Bounded Nature of Sigmoid acts as Implicit Regularization.
- Feature extraction: Only one CNN architecture was used.
- Weight decay: Little overfitting to regularize, only useful on larger networks.

Generalizability / Reliability:
- Low seed count. High variance.
- Only a single fixed out of sample dataset.
- Bottleneck: possible reason for no impact for network size and division?

Hard to derive strong conclusions.

Problem Complexity:



A solution: Decreasing Granularity (But exponentially decreases potential)

Proper Reward Signal?

$$G(\pi_A) > G(\pi_B) \iff \log \frac{E^A_{T-1}}{E_0} > \log \frac{E^B_{T-1}}{E_0}$$

Holds for PB Rewards!

7. Conclusion
Goal: Impact of FAM on an RL agents trading performance in EUR/USD.
Outcome: No (reliable) profit on unseen data, but FA components have significant impact.
Critical Tradeoff: learn vs generalize.
Example: ReLU vs Sigmoid

Other contributions include:
- Flexible, realistic, single-pair Gymnasium environment.
- Novel DP Formulation and application to Potential Based Rewards.

8. Future Work

1. Test hypotheses in discussion. Ex. More CNNs, test effects of bigger networks and more features.
2. Address Reliability Concerns.
3. Combine most successful elements.

Alternative Avenue:
Supervised Learning + DP Optimal as ground truth. Offers stability, directly and through shuffling.

References

TU Delft

Robert Mertens
N. Yorke-Smith
A. Papapantoleon
A. Kolarijani

As part of the course
Research Project
(CSE3000) 2025 Q4