

Prioritized Replay and Non-IID Sampling for Efficient Chess Evaluation Network Training

Lewis

Supervisor: Maximilien Gadouleau

October 21, 2025

Contents

1	Project Plan	2
1.1	Project Details	2
1.2	Project Description	2
1.3	Aims and Objectives	2
1.4	Preliminary Preparation	3
1.5	Deliverables	3
1.5.1	Basic Deliverables	3
1.5.2	Intermediate Deliverables	3
1.5.3	Advanced Deliverables	3
1.6	Timeline	4
1.7	References	4
2	Literature Survey	5
2.1	Introduction to Sample Efficiency in Chess Evaluation Networks	5
2.2	Key Themes in Sample-Efficient Chess Training	5
2.2.1	Prioritized Replay and Non-IID Sampling	5
2.2.2	Sample Weighting Functions	6
2.2.3	Information-Theoretic Training Objectives	6
2.2.4	Curriculum Learning	6
2.3	Assessment and Relation to Project	6
3	Critical Comparison with ChatGPT	7
3.1	Prompts Used	7
3.2	AI-Generated Literature Survey	7
3.3	Observations on Quality and Accuracy	7

1 Project Plan

1.1 Project Details

Student Name: Lewis

Supervisor Name: Maximilien Gadouleau

Project Title: Manifold-Aware Sampling for Efficient Chess Evaluation Network Training

1.2 Project Description

This project will investigate prioritized replay and non-IID sampling techniques to improve the sample efficiency of chess evaluation network training. Traditional approaches sample training positions uniformly from game databases, but this fails to account for the fact that not all positions provide equal learning value. Using Google DeepMind’s Chess-Bench dataset (10 million chess games with 15 billion Stockfish-annotated positions), I will implement adaptive buffers that replay high-weighted positions more frequently, creating non-stationary training distributions that focus on challenging examples and break from traditional SGD assumptions Schaul et al. [2016].

A key innovation will be the development of **prioritized replay buffers** and **non-iid sampling strategies** that dynamically prioritize positions based on their information content. Instead of treating all training examples equally, the system will adaptively sample high-value examples more frequently, allowing the training process to focus on currently challenging positions while maintaining a balance with exploration.

The project will begin with *naive position difficulty scoring* — simple heuristics for estimating training example importance — before progressing to more sophisticated weighting functions integrated into the replay mechanism. This approach will allow systematic comparison of different weighting schemes within the prioritized replay framework.

1.3 Aims and Objectives

The primary aim is to develop and validate prioritized replay and non-IID sampling techniques for chess evaluation network training that improve sample efficiency compared to random sampling. Key objectives include:

- Implement prioritized replay buffers that adaptively sample high-information positions more frequently Schaul et al. [2016], Mnih et al. [2015]
- Design non-iid sampling strategies that break from uniform distribution assumptions and create non-stationary training distributions (importance sampling background: Rubinstein and Kroese [2007], Owen [2013])
- Develop and compare different sample weighting functions integrated with replay mechanisms, starting with naive position difficulty scoring
- Progress from simple heuristics to sophisticated information-theoretic measures within the prioritized replay framework (influence functions: Koh and Liang [2017])
- Evaluate improvements in training speed and evaluation accuracy through controlled experiments comparing uniform vs. prioritized sampling

1.4 Preliminary Preparation

Before commencing the main implementation, I need to:

- Acquire datasets of labelled chess positions
- Understand information-theoretic measures in the context of neural network training (representation learning background: Kingma and Welling [2014], Rezende and Mohamed [2015])
- Review existing approaches to sample-efficient training and curriculum learning Bengio et al. [2009]

1.5 Deliverables

1.5.1 Basic Deliverables

- Chess position dataset and preprocessing pipeline using Google DeepMind’s Chess-Bench dataset (10 million games, 15 billion annotated positions)
- Basic HalfKP implementation for baseline testing
- Chess engine integration for playing strength evaluation and position generation
- Working evaluation network training loop with standard uniform sampling
- Naive position difficulty scoring functions integrated with prioritized replay buffers

1.5.2 Intermediate Deliverables

- Implementation of prioritized replay buffers with adaptive sampling based on position weights
- Comparison of multiple sample weighting functions within the replay framework (naive vs information-theoretic)
- Non-iid sampling strategies that create non-stationary training distributions
- Integration of dynamic weighting that adapts during training based on model performance
- Preliminary evaluation of sample efficiency improvements from prioritized vs uniform sampling

1.5.3 Advanced Deliverables

- Advanced weighting functions combining multiple information scores for optimal replay prioritization
- Dynamic replay mechanisms that adjust sampling distributions based on training progress
- Comprehensive ablation studies of different weighting functions and replay strategies

- Full integration of prioritized replay and non-iid sampling for end-to-end efficient training

1.6 Timeline

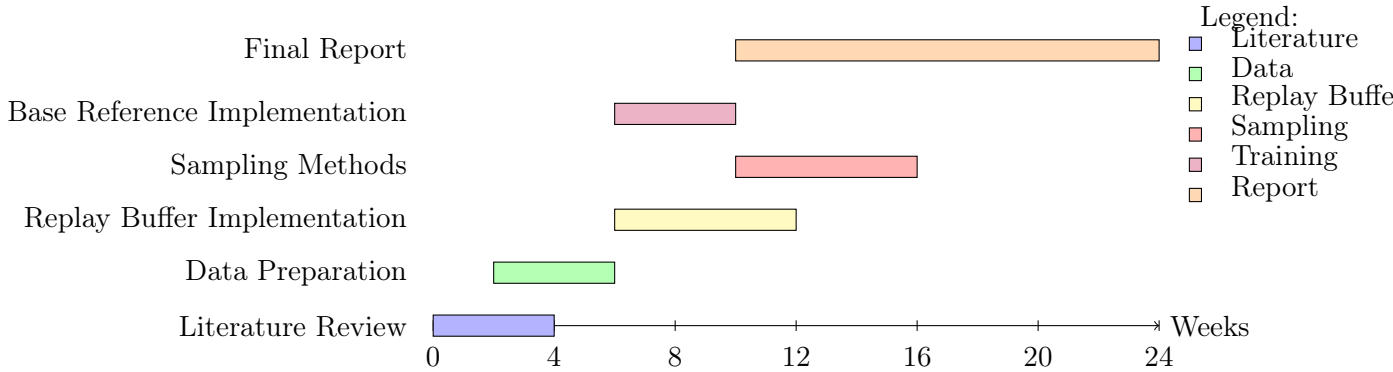


Figure 1: Project Timeline Gantt Chart

1.7 References

References

- Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, pages 41–48, 2009.
- Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations (ICLR)*, 2014.
- Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning (ICML)*, pages 1885–1894, 2017.
- Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- Art B. Owen. *Monte Carlo Theory, Methods and Examples*. Stanford University, 2013.
- Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning (ICML)*, pages 1530–1538, 2015.
- Reuven Y. Rubinstein and Dirk P. Kroese. *Simulation and the Monte Carlo Method*. Wiley, 2007.
- Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In *International Conference on Learning Representations (ICLR)*, 2016.

2 Literature Survey

2.1 Introduction to Sample Efficiency in Chess Evaluation Networks

Chess evaluation networks have become essential components of modern chess engines, but their training efficiency remains a significant challenge. Traditional approaches sample training positions uniformly from game databases, but this fails to account for the fact that not all positions provide equal learning value. Some positions are more challenging and informative for the model, requiring prioritized attention during training.

This survey examines prioritized replay and non-IID sampling as key strategies for sample-efficient training of chess evaluation networks. Instead of uniform sampling under IID assumptions, these techniques use adaptive buffers that replay high-weighted positions more frequently, creating non-stationary training distributions that focus on challenging examples and break from traditional SGD assumptions. The field bridges reinforcement learning techniques with the unique challenges of combinatorial game positions.

Key terms include:

- **Prioritized Replay:** Adaptive sampling that replays high-value examples more frequently
- **Non-IID Sampling:** Breaking from independent and identically distributed assumptions
- **Sample Weighting:** Functions that assign importance scores to training examples for replay prioritization
- **Naive Difficulty Scoring:** Simple heuristics for estimating example importance
- **Non-Stationary Distributions:** Training distributions that evolve and adapt during learning

2.2 Key Themes in Sample-Efficient Chess Training

2.2.1 Prioritized Replay and Non-IID Sampling

A key innovation in sample-efficient training is **prioritized replay**, where high-information positions are sampled more frequently than low-information ones, creating non-stationary training distributions. This breaks from the iid assumptions of traditional stochastic gradient descent and allows the training process to adaptively focus on currently challenging examples. Techniques include experience replay buffers with priority weighting and curriculum-based sampling that evolves the training distribution over time.

The core mechanism involves maintaining a replay buffer where positions are stored with associated weights. During training, positions are sampled proportionally to their weights, ensuring that high-value examples are revisited more often. This creates non-iid sampling patterns that adapt to the model’s learning progress, prioritizing positions that currently provide the most learning signal.

Key challenges include designing appropriate weighting functions, managing buffer size and update frequency, and preventing overfitting to high-weighted examples. Advanced implementations incorporate dynamic weighting that adjusts based on training progress and ensemble disagreement measures.

2.2.2 Sample Weighting Functions

A critical aspect of prioritized replay is the design of weighting functions that estimate the importance of training examples. *Naive difficulty scoring* provides a starting point with simple heuristics like material imbalance or position complexity. More sophisticated approaches use information-theoretic measures such as gradient norms, ensemble disagreement, and predictive uncertainty. The key challenge is developing weighting functions that correlate well with actual learning value while being computationally tractable for frequent replay buffer updates.

2.2.3 Information-Theoretic Training Objectives

Research has explored various measures to quantify the informativeness of training positions within replay frameworks. Gradient-based scores measure how much a position affects network parameters, while ensemble disagreement identifies positions where different models make conflicting predictions. These scores help prioritize replay on positions that maximize learning progress and adapt the non-iid sampling distribution.

2.2.4 Curriculum Learning

Curriculum learning suggests training on progressively more complex positions. Within prioritized replay, curriculum strategies help evolve the sampling distribution over time, starting with uniform sampling and gradually shifting to more focused non-iid patterns.

2.3 Assessment and Relation to Project

The literature reveals that sample efficiency in chess evaluation networks remains an open problem, with traditional uniform sampling far from optimal. While prioritized replay and non-IID sampling show significant promise for breaking from iid assumptions, their integration into coherent training strategies for chess is underexplored. My project will address this gap by developing prioritized replay buffers and non-IID sampling techniques that adaptively focus on challenging positions, with particular emphasis on dynamic weighting functions and non-stationary training distributions.

The key insight is that chess training can benefit greatly from non-iid sampling patterns that prioritize high-information positions through frequent replay. By implementing adaptive buffers that evolve the training distribution based on model progress, I will create training systems that break from traditional SGD assumptions and accelerate learning. This will build on recent advances in experience replay but apply them specifically to the combinatorial structure of chess through innovative weighting and sampling mechanisms.

Success will contribute to making strong chess evaluation more accessible through improved training efficiency, with potential applications to other domains requiring adaptive, non-stationary sampling from complex datasets.

3 Critical Comparison with ChatGPT

3.1 Prompts Used

Initial Prompt:

"Write a literature survey on sample-efficient training methods for neural networks, focusing on different sample weighting functions and importance sampling techniques."

Follow-up Prompt:

"Expand on naive difficulty scoring and simple heuristics for estimating training example importance, and discuss how these compare to more sophisticated information-theoretic approaches in practice."

3.2 AI-Generated Literature Survey

3.3 Observations on Quality and Accuracy