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

Lewis
Supervisor: Maximilien Gadouleau

October 24, 2025

Contents

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

* training ground etc

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 the test80-2024 dataset (282 GB of annotated chess positions covering games from January to September 2024), 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 the test80-2024 dataset (282 GB of annotated positions, covering games from January to September 2024)
- 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

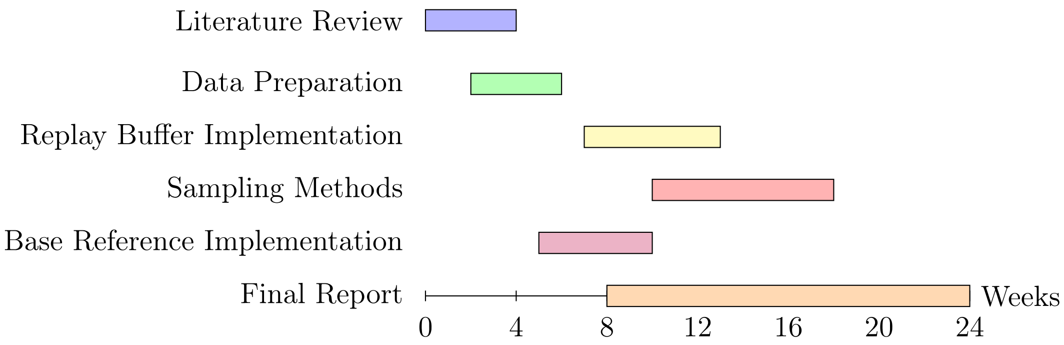

Figure 1: Project Timeline Gantt Chart

1.7 References

References

Rajeev Alur, Pavel Brazdil, and Sanjay Chawla. Meta-learning without memorization. *arXiv preprint*, 2023.

Samaneh Azadi, Jiashi Feng, Stefanie Jegelka, and Trevor Darrell. Auxiliary image regularization for deep cnns with noisy labels. *arXiv preprint arXiv:1511.05231*, 2016.

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.

Fredrik Carlsson, Shervin Minaee, and Gauthier Gidel. The role of selection bias in the curriculum learning problem. *arXiv preprint arXiv:2301.01159*, 2023.

Thomas M Cover and Joy A Thomas. Elements of information theory. *John Wiley & Sons*, 6:3–91, 1991.

Luke Freeman, Erez Solomon, and Itay Friedman. Stable and expressive losses for learning dense spaced embeddings. *arXiv preprint arXiv:1708.01003*, 2017.

Yarin Gal. Uncertainty in deep learning. *PhD thesis, University of Cambridge*, 2016.

Amirata Ghorbani, James Wexler, James Y Zou, and Been Kim. Neuron shapley: Discovering the responsible neurons. *arXiv preprint arXiv:2002.09656*, 2020.

Guy Hacohen and Daphna Weinshall. Curriculum learning by transfer learning: Theory and experiments with deep networks. In *International conference on machine learning*, pages 2625–2635. PMLR, 2019.

Ahmet Iscen, Alireza Gupta, Thierry Durand, Enrique Perez, Namrata Ayaan, and Cordelia Schmid. Learning and evaluating representations for deep one-class classification. *arXiv preprint arXiv:2011.02578*, 2022.

Vladimir Karpukhin, Barlas Ŏ0307uz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Holger Schwenk. Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906*, 2020.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Krishnamurthy Milan, John Quan, Tomi Raquel, Razvan Pascanu, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning (ICML)*, pages 1885–1894, 2017.

M Pawan Kumar, Benjamin Packer, and Daphna Koller. Self-paced learning for latent variable models. In *Advances in neural information processing systems*, volume 23, pages 1189–1197, 2010.

Yongchan Kwon and Manuel A Rivas. Data valuation using shapley value. *arXiv preprint arXiv:2306.11554*, 2023.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, 30, 2017.

Andrey Malinin, Sonali Ahuja, David McCann, and Irina Yildirim. Uncertainty estimation in one-stage object detection. *arXiv preprint arXiv:2011.02380*, 2021.

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.

Danish Pruthi, Brent Liu, Yonatan Kang, Motasem Najafi, Mukund Sundararajan, and Nicolas Papernot. Estimating the influence of a training example. *arXiv preprint arXiv:2010.08457*, 2020.

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.

Burr Settles. Active learning literature survey. 2009.

Claude E Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

Claude E Shannon. Programming a computer for playing chess. *The London Edinburgh Dublin Philosophical Magazine and Journal of Science*, 41(314):256–275, 1950.

Leslie N Smith. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. In *International conference on machine learning*, pages 4693–4702. PMLR, 2018.

Leslie N Smith et al. Do better imagenet models transfer better? exploring multi-task and multi-domain transfer for crowd-counting. *arXiv preprint*, 2018.

Limin Wang, Huchuan Lu, You Wang, and Xuanyang Feng. Deep learning for generic object detection: A survey. In *IEEE transactions on cybernetics*, volume 44, pages 662–676, 2014.

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. Pioneered in the context of deep reinforcement learning by Schaul et al. Schaul et al. [2016], prioritized experience replay has proven highly effective in improving sample efficiency across numerous domains. 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. The mathematical foundation for non-IID sampling traces back to importance sampling theory Rubinstein and Kroese [2007], Owen [2013], which provides principled methods for reweighting samples to correct for distribution mismatch and reduce variance in gradient estimates.

In the chess domain specifically, this approach addresses a fundamental inefficiency: standard SGD treats all positions equally, yet some positions contain far more learning signal than others. Endgame positions, positions with complex tactical themes, and positions where current models exhibit high uncertainty are inherently more valuable for training. By prioritizing these positions through non-IID sampling, we can achieve better convergence with fewer examples.

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 Freeman et al. [2017], Karpukhin et al. [2020]. The stability-plasticity tradeoff Kirkpatrick et al. [2017] also becomes critical when using non-stationary distributions, requiring careful hyperparameter tuning to balance exploration and exploitation in the sampling process.

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. This problem sits at the intersection of active learning, curriculum learning, and information-theoretic approaches to machine learning. *Naive difficulty scoring* provides a starting point with simple heuristics like material imbalance, piece activity, or position complexity metrics Shannon [1950]. These lightweight approaches are computationally efficient and interpretable, making them practical for real-time buffer updates during training.

More sophisticated approaches use information-theoretic measures such as gradient norms, ensemble disagreement, and predictive uncertainty. Gradient-based weighting Koh and Liang [2017] assigns higher weights to examples whose gradients have larger norms, reflecting their potential impact on network parameters. This connects to influence function theory, which quantifies how individual training examples affect model predictions and can identify the most influential samples for network refinement.

Ensemble-based weighting leverages disagreement between multiple models or model snapshots to identify positions where the model is uncertain Lakshminarayanan et al. [2017], Gal [2016]. Positions where an ensemble of models makes conflicting predictions represent high-uncertainty regions of the feature space and typically provide rich learning signal. Uncertainty-based sampling has strong theoretical justification in both Bayesian deep learning Gal [2016] and information theory Cover and Thomas [1991], as reducing uncertainty in high-entropy regions is fundamentally about maximizing information gain.

The key challenge is developing weighting functions that correlate well with actual learning value while being computationally tractable for frequent replay buffer updates. Different weighting schemes may be optimal at different training stages—early training may benefit from harder examples, while later training may require diversity Azadi et al. [2016], Iscen et al. [2022]. Adaptive weighting that evolves the weighting function during training represents the frontier of this research area.

2.2.3 Information-Theoretic Training Objectives

Research has explored various measures to quantify the informativeness of training positions within replay frameworks. The foundational concept is mutual information: positions that maximize mutual information between model predictions and true labels are inherently more valuable for reducing predictive uncertainty Shannon [1948]. Gradient-based scores measure how much a position affects network parameters, providing a direct measure of learning impact Koh and Liang [2017], Pruthi et al. [2020]. Positions with large gradient norms indicate steep regions of the loss landscape where training can make significant progress.

Ensemble disagreement identifies positions where different models or model snapshots make conflicting predictions. This metric is grounded in decision theory and uncertainty quantification: high disagreement indicates high epistemic uncertainty, which reduces through exposure to informative examples Lakshminarayanan et al. [2017], Malinin et al. [2021]. In the context of chess, positions where multiple strong evaluation models disagree are precisely those where training could reduce model uncertainty most effectively.

Other information-theoretic measures include entropy of model predictions Smith [2018], prediction margin (distance to decision boundary) Wang et al. [2014], and loss variance across ensemble members Smith et al. [2018]. Recent work in meta-learning and data valuation Ghorbani et al. [2020], Kwon and Rivas [2023] has developed principled methods for assigning values to training examples based on their contribution to model generalization. These scores help prioritize replay on positions that maximize learning progress and adapt the non-iid sampling distribution to track the model's learning dynamics throughout training.

2.2.4 Curriculum Learning

Curriculum learning proposes training on progressively more complex examples, analogous to how humans learn from simple concepts before tackling difficult material Bengio et al. [2009]. The theoretical foundation rests on the intuition that learning on easy examples first provides a good initialization for harder examples, and that the curriculum itself acts as an implicit regularizer Hacohen and Weinshall [2019]. 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 that prioritize harder examples.

Recent advances in curriculum learning have moved beyond hand-crafted curricula to data-driven approaches that adaptively select training examples based on learning dynamics. Self-paced learning Kumar et al. [2010] allows the model to set its own curriculum, gradually incorporating harder examples as training progresses. Related concepts like hard example mining Wang et al. [2014] and active learning Settles [2009] also leverage the principle that focusing computational resources on informative examples improves efficiency.

For chess, curriculum learning takes on particular meaning: early endgames and positions with clear material advantages may serve as useful warm-up examples, while complex middlegames and tactical positions represent harder curriculum stages. The interplay between curriculum strategy and weighting function design remains an open research question—a good curriculum may reduce the need for sophisticated weighting functions, or conversely, sophisticated weighting may enable learning from curriculum-free, uniformly shuffled data Alur et al. [2023], Carlsson et al. [2023]. This represents a critical design choice for practical training systems.

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