Artificial Intelligence Review (2024) 57:187
https://doi.org/10.1007/s10462-024-10813-3

Strategic data navigation: information value-based sample selection

Csanád L. Balogh[1,3] · **Bálint Pelenczei**[2] · **Bálint Kővári**[1,3] · **Tamás Bécsi**[1]

Accepted: 28 May 2024 / Published online: 25 June 2024

Abstract
Artificial Intelligence represents a rapidly expanding domain, with several industrial applications demonstrating its superiority over traditional techniques. Despite numerous advancements within the subfield of Machine Learning, it encounters persistent challenges, highlighting the importance of ongoing research efforts. Among its primary branches, this study delves into two categories, being Supervised and Reinforcement Learning, particularly addressing the common issue of data selection for training. The inherent variability in informational content among data points is apparent, wherein certain samples offer more valuable information to the neural network than others. However, evaluating the significance of various data points remains a non-trivial task, generating the need for a robust method to effectively prioritize samples. Drawing inspiration from Reinforcement Learning principles, this paper introduces a novel sample prioritization approach, applied to Supervised Learning scenarios, aimed at enhancing classification accuracy through strategic data navigation, while exploring the boundary between Reinforcement and Supervised Learning techniques. We provide a comprehensive description of our methodology while revealing the identification of an optimal prioritization balance and demonstrating its beneficial impact on model performance. Although classification accuracy serves as the primary validation metric, the concept of information density-based prioritization encompasses wider applicability. Additionally, the paper investigates parallels and distinctions between Reinforcement and Supervised Learning methods, declaring that the foundational principle is equally relevant, hence completely adaptable to Supervised Learning with appropriate adjustments due to different learning frameworks. The project page and source code are available at: https://csanadlb.github.io/sl_prioritized_sampling/.

Keywords Supervised learning · Classification · Sampling efficiency · Sample prioritization · Reinforcement learning

1 Introduction

Machine Learning has emerged as a rapidly evolving field, characterized by continual advancements and innovations. However, amidst plenty of state-of-the-art data-driven solutions addressing industrial challenges, a persistent difficulty exists: the inefficiency

Extended author information available on the last page of the article

of data sampling. This inefficiency arises from the inherent difficulty in determining the informational value of specific data points during the training process. Consequently, the optimization of sampling efficiency remains an unresolved issue, presenting a significant challenge for further advancement in Machine Learning techniques.

The inherent complexity in determining the relative importance of individual data samples merits further exploration. These samples vary in their contribution to the network's training process, with some holding more substantial pieces of new information or being more helpful in tuning the network weights. While prior determination, perhaps with the help of classic feature extraction, could offer preliminary insights into the significance of samples within the training dataset, it would ignore an important reality. The value of a sample is not solely reliant on the information within the raw data; it is also influenced by the current state of the neural network during training. Therefore, a dynamic approach for assessing the value of samples is essential, that continuously evaluates their significance in the context of the evolving state of the network. This paper introduces an innovative approach for dynamic prioritization of training data to enhance sampling efficiency in Machine Learning models. The primary objective is to develop a reliable metric capable of quantifying the informational value gained by a neural network from a specific sample. This approach aims to utilize samples rich in valuable information more frequently while reducing reliance on less informative ones.

Our methodology draws inspiration from existing prioritization strategies employed in Reinforcement Learning. In Reinforcement Learning, one often faces a vast number of potential experiences (training data points), making it impractical to retain all points in the memory buffer. Thus, a decision has to be made, which experiences to retain and which to discard, revealing a more apparent need for prioritization, and as such, several methods have already been established in this domain. Nevertheless, data prioritization remains relatively unexplored in Supervised Learning contexts. Although data points are typically static in these scenarios, hence removing the challenge of selection, sampling is often performed stochastically, relying on a uniform distribution without questioning the adequacy of assigning equal value to every data point. Since the informational value of different data points can greatly vary, a well-informed approach for prioritization could potentially yield superior outcomes compared to traditional sampling methods. Additionally, efficient sampling may potentially improve convergence characteristics or final performance metrics, such as accuracy of the network. This paper explores this premise, seeking to bridge the gap in data prioritization strategies between Reinforcement and Supervised Learning paradigms. The study is centered on the field of computer vision, specifically addressing the challenge of image classification. We demonstrate our concept using two widely recognized benchmark datasets for image classification and apply it across three well-known neural network architectures to prioritize among images.

1.1 Related work

Deep Learning proved to be an outstanding tool for computer vision applications, as neural network (NN) models are able to handle complex real-life visual information, achieving high accuracy on various well-known benchmark datasets (Voulodimos et al. 2018), such as CIFAR100 or ImageNet.

Several different approaches exist, however, sampling efficiency is rarely scrutinized. In Nguyen et al. (2011), a medical application is presented, where auxiliary label information is utilized to augment the information content of training data. The authors have proposed

a strategy, that involves leveraging supplementary probabilistic data indicating the confidence level associated with each label.

Prioritized sampling is particularly relevant in scenarios of class imbalance according to Dablain et al. (2023). In these cases, the training dataset contains a disproportionate amount of training samples for different labels. Techniques, such as oversampling the underrepresented data can enhance performance. However, the static nature of oversampling persists, as addressing class imbalance typically involves solely considering the number of data points in different classes.

The data prioritization method presented in this paper resembles the AdaBoost methodologies, initially introduced in Freund and Schapire (2002). These methods aim to construct an arbitrarily accurate strong predictor by combining multiple weak learners, each slightly outperforming random guessing. This process involves the re-weighting of training samples following the training of a weak learner, as visualized in Fig. 1. For more details, refer to: Hastie et al. (2009); Schapire (2013).

Active learning, akin to prioritization, enhances the efficiency of data annotation by allowing the model to select the samples that should be annotated. Sampling can focus on the most uncertain data points where the model struggles to make clear distinctions. The loss prediction module proposed by the authors in Yoo and Kweon (2019) is an innovative way to measure uncertainty by estimating potential losses for unlabeled inputs, not just enhancing the efficiency of data annotation, but offering a more universally applicable solution. Article (Beluch et al. 2018) compares the efficacy of different uncertainty estimation methods and acquisition functions with CNNs for image classification tasks. In Haut et al. (2018), the authors use active learning to tackle the time-consuming task of gathering hyperspectral images with the Bayesian-convolutional neural networks. The BALD (Bayesian Active Learning by Disagreement) algorithm, used by Houlsby et al. (2011) and later (Kirsch et al. 2019), enhances active learning by selecting data points that maximize the mutual information between the model's predictions and its parameters by identifying samples where the model's predictions are most uncertain.

To contextualize the methodology proposed in this study, it is essential to delve into the realm of Reinforcement Learning. In terms of training data prioritization, Prioritized Experience replay (Schaul et al. 2016) stands as a notable breakthrough. Its novel approach lies in the method of selecting experiences for replay based on their expected learning progress. Numerous adaptations of the original PER exist. For instance, Horgan et al. (2018) introduced a distributed architecture, that separates the processes of acting and learning. Another study in Brittain et al. (2020) has introduced Prioritized Sequence Experience Replay, an extension of the conventional PER, operating on entire sequences of transitions

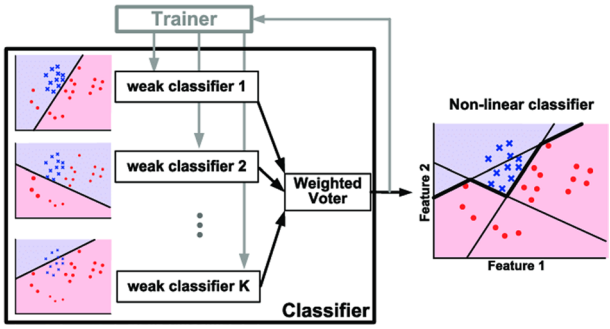

Fig. 1 Illustration of AdaBoost algorithm (Wang et al. 2015)

This study draws inspiration from an enhanced version of PER algorithm, proposed in Kővári et al. (2023). The methodology improves convergence speed and final state values by introducing an exploration element in the prioritization metric. PER inherently suffers from overfitting in certain situations, and the exploration term mitigates this risk while containing a tuning constant, that allows fine-tuning of the exploration–exploitation trade-off of the prioritization process. A trade-off that arises in several fields [e.g. Cuevas et al. (2014)]. This is equivalent to the overfitting–underfitting problem in supervised applications.

1.2 Contribution

Undoubtedly, recent years have witnessed significant progress in various aspects of Supervised Learning, including advancements in loss functions, neural network architectures, training algorithms, and data augmentation techniques. However, training sample prioritization, being a fundamental challenge in Reinforcement Learning rooted in the exploration-exploitation trade-off, remains a research gap in this domain. Consequently, stochastic sampling via a uniform distribution persists as the predominant technique for handling training data in this field.

In order to address the gap, this paper presents a novel approach to sampling prioritization. Our strategy efficiently navigates the training dataset to identify samples rich in new information by integrating insights from Reinforcement Learning into Supervised Learning applications. Moreover, the proposed methodology incorporates a sophisticated approach to modulate the inherent risk of overfitting associated with prioritization attempts, utilizing a dual-component metric. Additionally, the effectiveness of our approach is demonstrated through experiments in image classification scenarios, utilizing benchmark datasets and neural network architectures popular in literature.

It is worth mentioning that other methods are also concerned with defining the importance of training samples from the aspect of model performance, such as Active Learning. However, Active Learning utilizes this concept to distill an optimal dataset for a given problem, which also increases the model performance, thanks to finding the essential samples. Still, the approach presented in this paper utilizes sample importance differently by helping the model make the most of the existing dataset by sampling the critical data points more frequently.

2 Background

2.1 Supervised learning

Supervised Learning, a principal methodology in the field of Machine Learning and Artificial Intelligence, involves leveraging labeled datasets to train algorithms. This technique relies on input-output pairs, where input data is associated with known outputs, alias labels. During training, the algorithm iteratively adjusts the weights of a neural network based on the disparities between model predictions and actual labels, evaluated using a loss function. This iterative procedure, analogously to the Reinforcement Learning workflow detailed in Sect. 2.2, continues until the model achieves optimal accuracy, typically assessed through cross-validation techniques.

Alone in the computer vision domain encompasses a wide range of tasks, including regression, classification, object detection, segmentation, and tracking. Innovations in this field often result from architectural improvements, as demonstrated by the developments in Swin Transformers (Liu et al. 2021) and ConvNeXt (Liu et al. 2022b). Progress can also emerge from advancements in data augmentation techniques, such as those introduced in CutMix (Yun et al. 2019) and AutoMix (Liu et al. 2022a), or through novel optimization approaches like Sharpness Aware Minimization (SAM) (Foret et al. 2020). An intriguing instance of novelty, is when different SL or ML disciplines intersect and inspire one another. A notable example is the adaptation of Masked Autoencoders, initially developed for NLP applications, but have been elegantly utilized for vision tasks as well (He et al. 2021).

2.2 Reinforcement learning

Reinforcement Learning (RL) has become a key method for solving complex control and optimization problems, proving effective in diverse areas such as vehicle control (Kővári et al. 2020), traffic systems (Koh et al. 2020), robotics (Yan et al. 2020), and tracking control (Luo et al. 2016). Unlike other machine learning approaches, RL does not require predefined labels. Instead, it generates training data through a series of interactions between the learning agent and the environment. In each cycle, the agent assesses the current state, takes an action, and observes the environment's reaction. This ongoing process allows the agent to develop strategies that optimize performance based on feedback, or rewards, which measure the efficacy of actions toward achieving specific goals. The agent seeks to achieve the highest cumulated reward:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{1}$$

Here, G_t represents the cumulative reward weighted by a discount factor γ, emphasizing the balance between immediate and future rewards. This balance is crucial as it shapes the agent's strategy by adjusting focus between short-term actions and long-term benefits.

2.3 Sample prioritization

The methodology outlined in this paper draws inspiration from principles of Reinforcement Learning, a connection elaborated further in the following section. The aim is to optimize the selection of training data to maximize information gain. This refinement is achieved by formulating a probability distribution for sample selection, which represents the weighting of information gain. Two distinct metrics are laid out, both yielding a sampling probability distribution reflective of the anticipated information gain expected from utilizing the specified samples.

2.3.1 Prioritized experience replay

Temporal Difference (TD) error, as introduced by Sutton (1988), serves as a crucial metric in value-based learning methodologies within Reinforcement Learning. It quantifies the gap between the expected and observed values of a state or state-action pair, computed from the immediate reward plus the discounted future value:

$$\delta_t = R_{t+1} + \gamma V(S_{t+1}) - V(S_t) \tag{2}$$

$$\delta_t = R_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t) \tag{3}$$

Here, δ_t represents the TD error at time step t, capturing the mismatch between predicted and actual outcomes, thus indicating potential learning value in experiences. This metric is crucial for Prioritized Experience Replay (PER), which utilizes TD error to guide the sampling of experiences, enhancing learning efficiency and model performance. In PER, introduced by Schaul et al. (2016), experiences are weighted based on their TD error, ensuring that those with higher errors are sampled more frequently to optimize learning outcomes. This approach is mathematically described by:

$$p(i) = |\delta_i| + \epsilon \tag{4}$$

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha} \tag{5}$$

In this model, $p(i)$ quantifies the priority level of sample i, with α modulating the extent of prioritization. This technique not only boosts convergence rates but also ensures a balanced representation of experiences, mitigating bias through controlled sampling.

This integration of TD error with PER supports a more strategic, informed approach to training in Reinforcement Learning, optimizing both the rate and quality of learning through prioritized experience sampling.

3 Strategic data navigation

The exploration–exploitation trade-off in Reinforcement Learning mirrors the challenges of underfitting and overfitting in Supervised Learning. Excessive exploration without sufficient refinement of learned states may lead to underfitting, where the agent fails to gather adequate information for effective policy development. Conversely, excessive exploitation can result in overfitting to a limited set of experiences, hindering the agent's ability to grasp the broader problem structure. Prioritization inherently tends toward exploitation by favoring certain samples over others. The Upper Confidence Bound strategy, proposed in Kővári et al. (2023), presents a sophisticated approach to finely balance exploitation and exploration, ensuring a more efficient learning process. The mathematical formulation of the UCB method is shown in Eq. 6 as:

$$\mathrm{UCB}_i = \frac{p_i^\alpha}{\sum_k p_k^\alpha} + c_p \cdot \sqrt{\frac{2 \cdot \ln(\max_k n_k)}{n_i + \epsilon}} \tag{6}$$

where UCB value of sample i is given as a summation of the exploration and the exploitation components. Exploitation remains the normalized priority value, consistent with the formulation introduced in Eq. 5. However, exploration is quantified by n_i, representing the fit count indicating how many times sample i has been utilized for network updates, relative to n_k, the maximum fit count among all experiences stored in the memory buffer. The constant ϵ is introduced to prevent division by zero, while c_p acts as a parameter, constrained within the interval of [0; 1], designed to finely tune the equilibrium between exploration

and exploitation. This delicate balance empowers the agent to navigate the learning environment's state space more efficiently, leveraging the strengths of both strategies to achieve optimal performance.

3.1 Probability-based approach

The concept of exploration is relatively straightforward; utilizing the frequency of a sample's usage in training as a metric helps to mitigate overfitting by avoiding excessive reliance on a limited set of samples, as seen in previous discussions. This principle is consistent across both Reinforcement Learning and Supervised Learning domains.

Nevertheless, delving into exploitation requires deeper investigation. Here, the counterparts of the temporal difference, or its constituent elements within Supervised Learning must be considered. Both methodologies involve a predicted value—one generated by the action network in case of RL and the other by the sole network in SL. However, unlike Reinforcement Learning, where a dedicated target value is employed, Supervised Learning relies on ground truth.

In this context, the closest parallel to temporal difference lies in the disparity between the ground truth probability and the network's confidence in its prediction, as expressed in Eq. 7:

$$\mathrm{PB}_i = y_{i,j} - \max_l \ \sigma(\hat{y}_i(x_i; l, \theta)) \tag{7}$$

where

$$j = \arg\max \ \hat{y}_i(x_i; l, \theta)$$

In the equation above, $\hat{y}_i$ represents the probabilistic model outputs for sample i, l denotes the set of labels, θ denotes parameters of the neural network, σ signifies the softmax function, $y_{i,j}$ denotes the ground truth for the j-th class of sample i, assuming a binary value of either 0 or 1 and x_i represents the input sample i. Essentially, the Probability Error metric quantifies the difference between the highest predicted probability value and the ground truth of the corresponding label.

This disparity in probabilities highlights the informational value of a training sample, explicitly indicating the extent of divergence between the network's prediction and the actual ground truth. The objective of the training process is to minimize this discrepancy between predicted outcomes and ground truth, with this metric designed precisely for this purpose. Further details on this metric is provided in Sect. 4.1.

In the proposed methodology, illustrated in Fig. 2, the training process initiates by establishing a uniform probability distribution, which forms the basis for stochastic sampling from the training dataset. Hence, the selection mechanism ensures a balanced representation from the onset. Following the step of sample selection, loss computation is performed for the currently selected batch of data. The subsequent step involves model fitting based on the calculated loss, enabling the acquisition of both the exploration and exploitation metrics. Thereafter, weights, aka UCB values, are computed. Finally, upon the weight assignment to each sample, the probability distribution undergoes an update based on these recalibrated weights. This update assigns higher probability of selection to samples anticipated to possess greater information density at a given time, thereby establishing a loop for dynamic weighting in the sampling procedure.

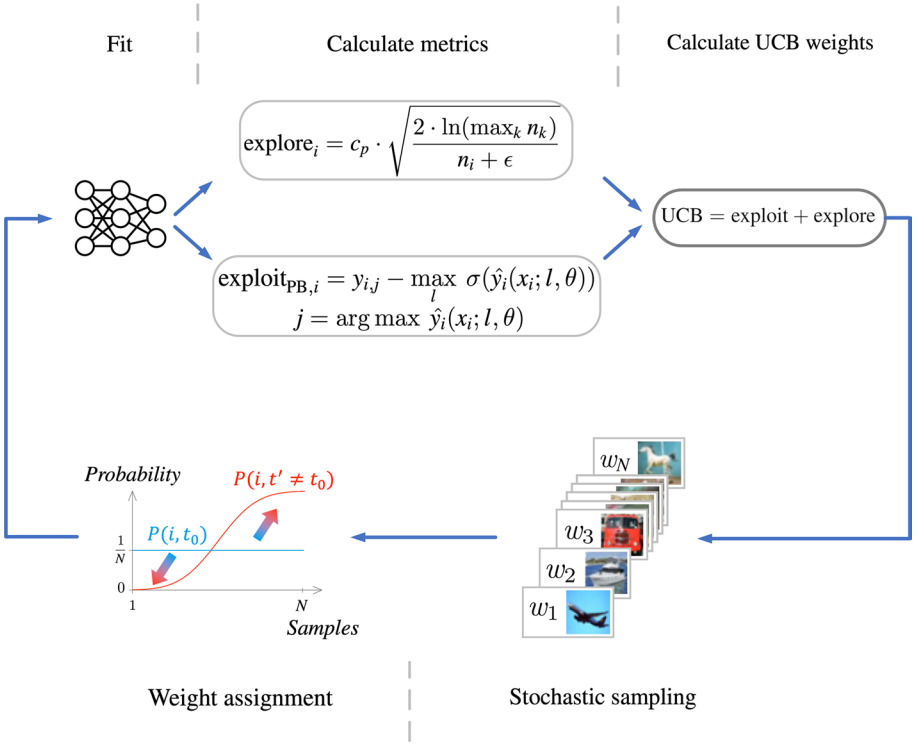

Fig. 2 Illustration of the prioritized sampling methodology. The sampling probability distribution is updated at each iteration of the training process

3.2 Abstract formalism

As previously highlighted, while the exploitation strategy requires adjustments compared to the RL framework, the fundamental aim remains to be the development of a metric, that evaluates the expected information gain from specific samples. The PB error, derived directly from the TD error, embodies a perspective rooted in Reinforcement Learning. In contrast, the Label Change Error offers a measure tailored for a Supervised Learning context. It is reasonable to argue, that the frequency of class changes for a training point indicates the network's challenge in identifying the content of the image. Such data points could offer valuable learning opportunities for the network due to their complexity. This leads to an additional metric, namely Label Change Error, providing a more aligned scheme with Supervised Learning. In this case, the exploitation element is replaced with a counter c_l registering how many times each sample changes its label. Subsequently, this label change count is utilized to compute the sample priorities, as shown in Eq. 8:

$$\text{LC}_i = c_{l,i} \tag{8}$$

The Label Change Error provides a simpler and more refined measure for the level of surprise while maintaining consistency with the main philosophy. Further details and advantages of this metric are demonstrated in Sect. 4. This metric is specifically designed for computer vision and classification problems, aiming to identify images that are particularly

beneficial for the training process. Applications outside this domain would necessitate a careful reevaluation of the underlying calculations. Initially, our approach leveraged a reinforcement learning-based solution, redesigned for image classification. Despite the distinct challenges posed by this setting-unlike selecting actions for an agent-a minimal and direct modification allows the transition from using prediction probabilities to employing label change frequency as a metric for information gain.

4 Experiments

4.1 Probability error

Both proposed metrics introduce an additional hyperparameter, c_p, crucial for maintaining an appropriate balance between underfitting and overfitting. Optimal tuning of c_p is vital for effective sample prioritization. The Probability Error, described in Sect. 3.1, shows promises in enhancing classification accuracy. However, despite meticulous selection of c_p values, the improvement in accuracy is relatively modest. A fundamental limitation arises from the debatable effectiveness of employing probability differences as a metric for error evaluation. This issue is exemplified by the CIFAR dataset, where images of willow, oak and maple trees exhibit either indistinguishable characteristics or pose significant identification challenges, even to trained observers. In order to get an understanding on the difficulty of the task, this phenomenon is illustrated in Fig. 3. In such cases, conventional labeling approaches dictate, that an image of a willow tree is classified as 0% oak, 0% maple and 100% willow, an oversimplification, that disregards shared characteristics among these tree species. As a result, a proficiently trained model might correctly identify a willow tree as such, although with low confidence due to the close resemblance to other tree classes, resulting in a substantial Probability Error. Expecting the model to predict with 100% certainty in favor of the willow tree to mitigate the error is unrealistic under these circumstances. Therefore, reliance on Probability Error as a metric prioritizes samples near class boundaries within the feature space, which ideally should be recognized as closely related. This tendency slightly encourages the network to overfit a few samples lacking unique informational content. Although the exploration term can partially neutralize this tendency, the underlying issue remains unresolved.

Another potential source of bias within the metric arises from instances, where identical images are assigned disparate labels. This phenomenon is exemplified by certain examples within the CIFAR100 dataset, as illustrated in Fig. 4. Given the close relationship between these categories, one might argue, based on prior reasoning, that classifying these images into different categories does not constitute a significant error. However, this scenario presents the network conflicting information, suggesting, that a single image could belong to multiple classes—an assertion such, that while plausible in specific contexts [e.g.

Fig. 3 Examples of cross-class similarity in CIFAR100 dataset (graphical quality is inherent for the CIFAR dataset)

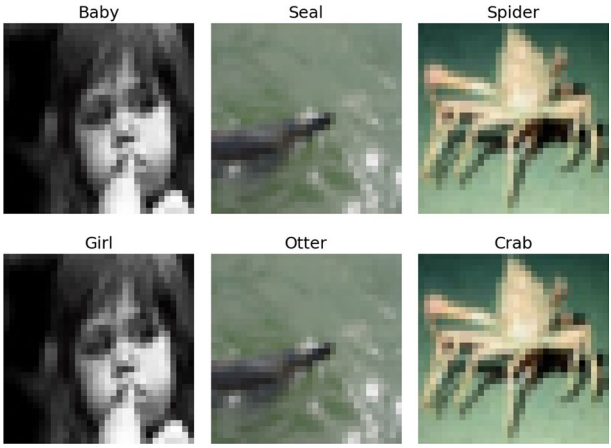

Fig. 4 Examples of duplicate labels with different classes assigned in CIFAR100 dataset

in case an image of a baby girl might reasonably fit into more than one category, invoking multi-label classification, as described by Fürnkranz et al. (2008)], still poses challenges. Viewing this perspective, it remains problematic, that the probability-based error metric for these images is elevated due to their positioning at the intersection of two classes. This situation underscores a limitation of the PB error metric: it penalizes the network for ambiguity inherent in the dataset itself, rather than inadequacies in the network's classification capabilities. This issue affects a small minority of all training data, but is present nonetheless and without efficient smoothing (i.e. tuning of the c_p parameter), it can be significant.

4.2 Label change error

Despite the limitations associated with the probability-based error metric, it has shown a modest improvement in accuracy indicating, that the core principle of prioritizing information gain remains valid. To address the shortcomings of the PB error metric, we investigated an alternative metric, termed Label Change Error. This novel approach successfully addresses the drawbacks of the PB error, while adhering to the original concept of emphasizing information gain. The label change metric leverages the behaviour of data points, that frequently alter their class affiliation, typically residing at class boundaries. By focusing on these critical data points, the metric aims to train the network to discern intricate patterns more effectively.

In contrast, the challenge of inherently low probabilities for certain data points, a notable issue with the PB error metric, becomes irrelevant under the label change approach. This method operates on a more abstract level, bypassing direct probability assessment. Consequently, data points firmly categorized within a specific class, even if with lower confidence due to their resemblance to other classes in the dataset, are not disproportionately emphasized. Instead, attention is directed towards data points presenting classification challenges, allowing the network to examine and learn from these cases with greater intensity. This strategic shift ensures, that learning is focused on areas, where the network can achieve the most significant gains in understanding the dataset's complexity, thereby enhancing overall model performance. It is worth noting, that maintaining a balance against overfitting remains crucial, emphasizing the importance of selecting the correct c_p value.

Table 1 Comparison of validation accuracy across different configurations on CIFAR datasets

	Dataset	ResNet50 (%)	MobilenetV-3Large (%)	EfficientNet_B1 (%)
Uniform sampling	CIFAR100	76.02	66.89	66.72
	CIFAR10	94.49	91.30	91.85
Prioritized sampling	CIFAR100	76.84	67.66	67.51
	CIFAR10	94.90	91.75	92.37
Accuracy gain	CIFAR100	+0.817	+0.773	+0.787
	CIFAR10	+0.408	+0.450	+0.513

Table 2 Comparison of validation accuracy across different configurations on TinyImagenet

	Dataset	ResNet50 (%)	MobilenetV-3Large (%)	EfficientNet_B1 (%)
Uniform sampling	TinyImagenet	63.84	57.75	57.22
Prioritized sampling	TinyImagenet	64.57	58.62	57.92
Accuracy gain	TinyImagenet	+0.731	+0.871	+0.702

5 Results

Throughout the experiments, the methodology was restricted to data augmentation and hyperparameter optimization, with no advanced modifications applied to the base neural network architectures. The principal objective of this study is to demonstrate the effectiveness of information gain-based prioritization in enhancing model efficiency and to showcase that the proposed metric successfully embodies this concept. The impact of prioritization is illustrated via the task of image classification. Tables 1 and 2 present the obtained validation accuracy gain, demonstrating the effectiveness of the approach. These accuracy metrics represent mean values derived from 6 distinct random seeds. This approach ensures that the study focuses on evaluating the influence of information gain-based prioritization on model performance, rather than the potential benefits of novel network architectures. By leveraging common data augmentation techniques such as Gaussian blur, color jitter, random resize crop, and random flips, we aim to simulate real-world variations in the dataset, thus providing a robust testing ground for the proposed prioritization methodology.

The convergence of validation accuracy of the CIFAR datasets is depicted in Fig. 5, for ResNet50 and Mobilenet V3. Similarly, Fig. 6 represents the accuracy of Mobilenet V3 and Efficientnet B1 on TinyImagenet. The noticeable increase in accuracy gain observed on the more complex CIFAR100 dataset indicates scalability. Furthermore, the consistent accuracy improvements demonstrate, that the proposed methodology reliably supports model performance across different network architectures, datasets, and random seeds. The proposed method has been tested on the CIFAR datasets, as mentioned earlier. The sample prioritization strategy also boosts the performance considerably on the TinyImagenet dataset, which features double the size and number of classes, thereby presenting a significantly higher level of complexity. As illustrated in the convergence plots of Fig. 6, the final validation accuracy consistently exceeds the baseline values. The approximate gain of 0.6% in

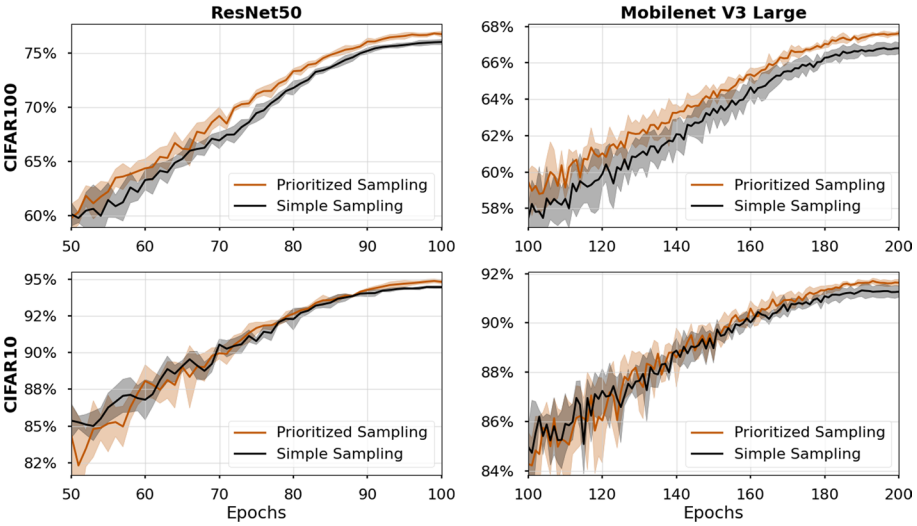

Fig. 5 Converge curve of validation accuracy with standard deviation bound from six distinct random seeds on CIFAR datasets

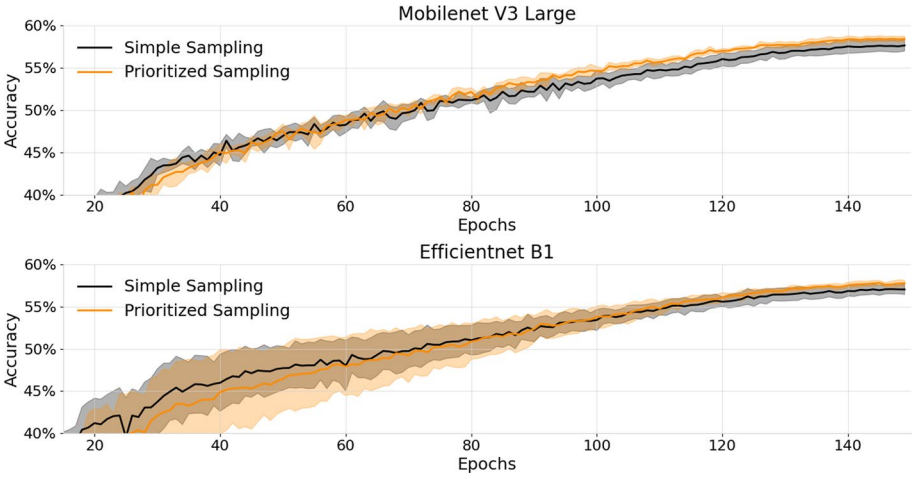

Fig. 6 Converge curve of validation accuracy with standard deviation bound from six distinct random seeds on TinyImagenet dataset

accuracy across various datasets substantiates the dataset-independent nature of the metric, demonstrating its capability to efficiently identify the most valuable samples across diverse datasets and classes.

In the context of Reinforcement Learning, prioritization benefits from extended periods to influence outcomes as agents often undergo thousands of episodes. It is notable that, in comparison, the impact of prioritization becomes evident within a considerably shorter time frame. This difference can be partly attributed to the dynamic nature of the Reinforcement Learning buffer, as opposed to the static nature of supervised datasets.

Upon examining the images assigned the highest weights, the issue of duplicate data is still observed, as outlined in 4.1, albeit to a lesser extent. Despite the occurrence of low probability values, labels often converge to a resting point. This convergence prevents the prioritization method from targeting these samples for focused learning. This observation indicates a reduction in the incidence of duplicate data problems and explains the success of the label change metric.

The observed effects of prioritization are significant, though not overwhelming, due to the exploration component of the proposed metric. For instance, in 100 epochs using the ResNet50 architecture, the most notable difference in the utilization frequency (fit count) of a sample reached 44, indicating that while samples with lower priorities are still extensively used, thus ensuring their informational content is leveraged, the mechanism also effectively emphasizes the selection of more crucial samples. To better understand this phenomenon, the t-SNE (t-Distributed Stochastic Neighbor Embedding) algorithm was employed. t-SNE is known for its ability to reduce the dimensionality of high-dimensional data, making it especially useful for visualizing such data in a low-dimensional space, by transforming similarities between data points into joint probabilities and then minimizing the Kullback-Leibler divergence between these probabilities across both high-dimensional and low-dimensional spaces. This process effectively groups similar data points together while separating dissimilar ones. The visualization provided in Fig. 7 demonstrates how t-SNE distinguishes the ten classes of the CIFAR10 dataset, represented with various colors. Here, the size of the points correlates with the frequency of a sample's use in model training. A larger point indicates a higher difference in utilization rate.

A critical conclusion from this study highlights the significance of the c_p parameter, which balances exploration and exploitation. Selecting an appropriate value for this parameter is crucial, as suboptimal choices may result in minimal to no beneficial effect. While the choice of parameter value appears to be slightly influenced by the neural network architecture, it is significantly affected by the dataset. Interestingly, our findings also suggest, that prioritization does not notably increase computational demands or training runtime.

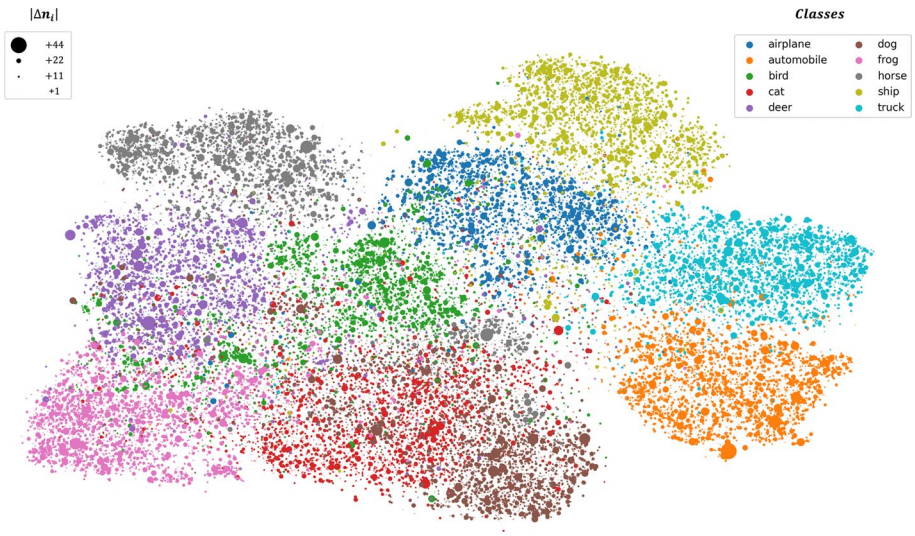

Fig. 7 Variability in the usage of data points, categorized by class through the t-SNE algorithm

6 Conclusion

This research addresses a fundamental challenge in the field of Machine Learning: the inefficiency of data sampling. Despite continuous advancements and innovations in the domain, the optimization of sampling efficiency remains an unresolved issue.

The essence of the problem lies in the inherent difficulty of determining relative importance of individual training samples during the training process. Traditional approaches fall short in providing a comprehensive solution, as they overlook the dynamic nature of a sample's value due to the evolving state of the neural network over training time.

To address this challenge, this paper proposes an innovative approach for dynamic prioritization of training data to enhance sampling efficiency in Supervised Learning models, while exploring the boundary between Reinforcement and Supervised Learning techniques in Machine Learning.

With the task of classification serving as a primary demonstration tool for our methodology's effectiveness, and RL being a source of inspiration, a vast number of existing prioritization methods are presented in Sect. 1.1, followed by a deeper dive into the RL framework in Sect. 2.2, highlighting the similarities and differences of these methods. Sections 3.1, 3.2 elucidate the formalization of the proposed metrics, while Sect. 4 describes practical considerations, conceptual risks and benefits. Although the same formalism cannot be directly applied due to the different nature of learning frameworks, the general utility of prioritization is demonstrated after the appropriate modifications through classification results in Sect. 5.

The widely recognized CIFAR100 and CIFAR10 datasets have been used for benchmarks, aiming to assess the overall impact of our methodology without employing special augmentations. Similarly, the original ResNet50 and Mobilenet V3 architectures are utilized without alterations. This study endeavors to emphasize the parallelism between different Machine Learning techniques through strategic data navigation, leveraging a concept that is generalizable across a diverse set of applications.

In our future endeavors, based on the potential of the developed sample prioritization method, our aim is to expand its application as this study has opened up numerous promising opportunities for future research. A natural extension involves experimenting with a broader array of datasets. This expansion would illuminate any dataset-specific nuances affecting the efficacy of prioritization. In particular, exploring datasets with varying levels of complexity, diversity and size would enable to deepen our understanding of how prioritization performs across different dataset characteristics.

Additionally, extending the application of prioritization to other tasks beyond image classification, such as object detection or semantic segmentation, is an exciting frontier. Object detection, with its unique challenges and requirements, could significantly benefit from prioritization strategies, especially in handling imbalanced datasets or focusing on rare, but critical objects. This exploration would necessitate adapting prioritization metrics and strategies, accounting for handling bounding boxes and multiple objects per image, for instance, as the current metric is designed to seek out important images. With multiple bounding boxes per image, the concept needs to be refined. Another promising avenue for future work involves exploring the integration of prioritization with different models.

Furthermore, a deeper integration of Reinforcement Learning and Supervised Learning presents another intriguing field for investigation. Specifically, Reinforcement Learning may presumably be utilized directly for the prioritization process through prediction of sample weights, which would introduce the possibility of a synergistic interaction between

the two approaches. An RL agent could be developed to operate together with the Supervised Learning agent and identify an optimal prioritization strategy. This approach would not only allow a dynamic and adaptive prioritization mechanism, but it would also yield to a direct and explicit cooperation between Supervised and Reinforcement Learning frameworks.

Author contributions Not applicable.

Funding Open access funding provided by Budapest University of Technology and Economics. This work was supported by the European Union within the framework of the National Laboratory for Autonomous Systems (RRF-2.3.1-21-2022-00002). The research reported in this paper is part of project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme. Tamás Bécsi was supported by BO/00233/21/6, János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Data availibility Project page and source codes are available at: https://csanadlb.github.io/sl_prioritized_sampling/

Materials availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no competing interests to declare relevant to this article's content.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Beluch WH, Genewein T, Nürnberger A, et al (2018) The power of ensembles for active learning in image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Brittain M, Bertram J, Yang X, et al (2020) Prioritized sequence experience replay. arXiv: 1905.12726

Cuevas E, Echavarría A, Ramírez-Ortegón MA (2014) An optimization algorithm inspired by the states of matter that improves the balance between exploration and exploitation. Appl Intell 40:256–272. https://doi.org/10.1007/s10489-013-0458-0

Dablain D, Krawczyk B, Chawla NV (2023) Deepsmote: fusing deep learning and smote for imbalanced data. IEEE Trans Neural Netw Learn Syst 34(9):6390–6404. https://doi.org/10.1109/TNNLS.2021.3136503
Foret P, Kleiner A, Mobahi H, et al (2020) Sharpness-aware minimization for efficiently improving generalization. arXiv preprint arXiv:2010.01412
Freund Y, Schapire R (2002) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. https://doi.org/10.1006/jcss.1997.1504
Fürnkranz J, Hüllermeier E, Loza Mencía E et al (2008) Multilabel classification via calibrated label ranking. Mach Learn 73:133–153. https://doi.org/10.1007/s10994-008-5064-8
Hastie T, Rosset S, Zhu J et al (2009) Multi-class adaboost. Stat Interface 2(3):349–360
Haut JM, Paoletti ME, Plaza J et al (2018) Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach. IEEE Trans Geosci Remote Sens 56(11):6440–6461. https://doi.org/10.1109/TGRS.2018.2838665
He K, Chen X, Xie S, et al (2021) Masked autoencoders are scalable vision learners. arXiv: 2111.06377
Horgan D, Quan J, Budden D, et al (2018) Distributed prioritized experience replay. arXiv: 1803.00933
Houlsby N, Huszár F, Ghahramani Z, et al (2011) Bayesian active learning for classification and preference learning. arXiv: 1112.5745
Kirsch A, van Amersfoort J, Gal Y (2019) Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. In: Wallach H, Larochelle H, Beygelzimer A, et al (eds) Advances in Neural Information Processing Systems, vol 32. Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2019/file/95323660ed2124450caaac2c46b5ed90-Paper.pdf
Koh S, Zhou B, Fang H et al (2020) Real-time deep reinforcement learning based vehicle routing and navigation. Appl Soft Comput. https://doi.org/10.1016/j.asoc.2020.106694
Kővári B, Hegedüs F, Bécsi T (2020) Design of a reinforcement learning-based lane keeping planning agent for automated vehicles. Appl Sci. https://doi.org/10.3390/app10207171
Kővári B, Pelenczei B, Bécsi T (2023) Enhanced experience prioritization: a novel upper confidence bound approach. IEEE Access 11:138488–138501. https://doi.org/10.1109/ACCESS.2023.3339248
Liu Z, Lin Y, Cao Y, et al (2021) Swin transformer: Hierarchical vision transformer using shifted windows. arXiv: 2103.14030
Liu Z, Li S, Wu D, et al (2022a) Automix: Unveiling the power of mixup for stronger classifiers. arXiv: 2103.13027
Liu Z, Mao H, Wu CY, et al (2022b) A convnet for the 2020s. arXiv: 2201.03545
Luo B, Liu D, Huang T et al (2016) Model-free optimal tracking control via critic-only q-learning. IEEE Trans Neural Netw Learn Syst 27(10):2134–2144. https://doi.org/10.1109/TNNLS.2016.2585520
Nguyen Q, Valizadegan H, Seybert A, et al (2011) Sample-efficient learning with auxiliary class-label information. AMIA Annu Symp Proc
Schapire RE (2013) Explaining adaboost. In: Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik. Springer, Charm. 37–52
Schaul T, Quan J, Antonoglou I, et al (2016) Prioritized experience replay. arXiv: 1511.05952
Sutton R (1988) Learning to predict by the method of temporal differences. Mach Learn 3:9–44. https://doi.org/10.1007/BF00115009
Voulodimos A, Doulamis N, Doulamis A et al (2018) Deep learning for computer vision: a brief review. Comput intell Neurosci. https://doi.org/10.1155/2018/7068349
Wang Z, Zhang J, Verma N (2015) Realizing low-energy classification systems by implementing matrix multiplication directly within an ADC. IEEE Trans Biomed Circuits Syst 9:1–1. https://doi.org/10.1109/TBCAS.2015.2500101
Yan C, Xiaojia X, Wang C (2020) Towards real-time path planning through deep reinforcement learning for a UAV in dynamic environments. J Intell Robot Syst. https://doi.org/10.1007/s10846-019-01073-3
Yoo D, Kweon IS (2019) Learning loss for active learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
Yun S, Han D, Oh SJ, et al (2019) Cutmix: regularization strategy to train strong classifiers with localizable features. arXiv: 1905.04899

Authors and Affiliations

Csanád L. Balogh[1,3] · Bálint Pelenczei[2] · Bálint Kővári[1,3] · Tamás Bécsi[1]

✉ Tamás Bécsi
becsi.tamas@kjk.bme.hu

Csanád L. Balogh
csanadlevente.balogh@edu.bme.hu

Bálint Pelenczei
pelenczei.balint@sztaki.hun-ren.hu

Bálint Kővári
kovari.balint@kjk.bme.hu

[1] Department of Control for Transportation and Vehicle Systems, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Budapest 1111, Hungary

[2] Systems and Control Laboratory, HUN-REN Institute for Computer Science and Control (SZTAKI), Budapest 1111, Hungary

[3] Asura Technologies Ltd., Budapest 1122, Hungary