



**CEAI**  
Center of Excellence in Artificial Intelligence

artiq

# ***Przyspieszanie treningu głębokich sieci neuronowych przy użyciu podejścia strumieniowego z mechanizmem wartości bazowych***

---

Mateusz Wojtulewicz

X Spotkanie grupy PL-SIGML

Kraków

20.11.2024

# Tradycyjne metody doboru danych

Metody wybierające dane uczące w sposób ustrukturyzowany, epokowy



**Możliwy problem:** Redundancja kroków uczących

# Nowe metody doboru danych

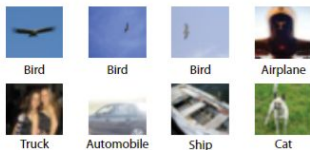
## Related Works

Metody zmieniające tradycyjną procedurę wyboru przykładów w trakcie uczenia

### Selective-Backprop



(a) Examples chosen least frequently by SB



(b) Examples chosen most frequently by SB

Jiang, Angela H., et al. "Accelerating deep learning by focusing on the biggest losers." (2019).

### KAKURENBO

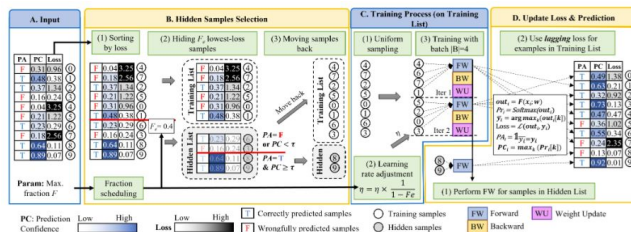


Figure 1: Overview of KAKURENBO. At each epoch, samples are filtered into two different subsets, the training list and the hidden list, based on their loss, prediction accuracy (PA), and prediction confidence (PC), with a maximum hidden fraction of  $F$ . PA and PC are used to drive sample move back decisions. Samples in the training list are processed using uniform sampling without replacement. The loss and the prediction accuracy, calculated from the training process, are reused to filter samples in the next epoch. For samples on the hidden list, KAKURENBO only calculates the loss and PA by performing the forward pass at the end of each epoch.

Nguyen, Trung Thao, et al. "KAKURENBO: adaptively hiding samples in deep neural network training." Neural Information Processing Systems Conference (2023).

### Podjęcie strumieniowe



Duda, Piotr, et al. "On training deep neural networks using a streaming approach." Journal of Artificial Intelligence and Soft Computing Research 10.1 (2020): 15-26.

# Zamysł podejścia strumieniowego

Duda, Piotr, et al. "On training deep neural networks using a streaming approach." Journal of Artificial Intelligence and Soft Computing Research 10.1 (2020): 15-26.

Podejście to rezygnuje z tradycyjnego epokowego doboru danych

Zamiast tego, tworzy strumień danych ze zbioru uczącego, stąd nazwa

## **Wyzwanie:**

Jak wybierać przykłady uczące do strumienia, aby przyspieszyć uczenie sieci?

## **Pomysł:**

Częściej wybierać przykłady trudniejsze

# Konstrukcja podejścia strumieniowego

**Pomysł:** Częściej wybierać przykłady trudniejsze

Zbiór uczący  $\rightarrow$  Populacja statystyczna

$$D = \{(x_i, d_i)\} \longrightarrow P_D = \{(x_i, d_i, p_i)\} \quad \sum_{i=1}^N p_i = 1$$

Wartości prawdopodobieństwa wyboru (ang. *Probability of Drawing, PoD*) tworzą rozkład prawdopodobieństwa nad zbiorem uczącym.

W założeniu: im trudniejszy przykład uczący, tym wyższa jego wartość PoD.

Strumień danych uczących

$$S_D = (s_1, s_2, \dots, s_t : s_j = (x_{i_j}, d_{i_j}), P(i_j = i) = p_i, i_j = 1, \dots, N, j = 1, \dots, t)$$

Strumień danych uczących to sekwencja przykładów uczących niezależnie próbkowanych z populacji  $P_D$  zgodnie z rozkładem prawdopodobieństwa zdefiniowanym przez wartości PoD.

# Algorytm uczenia strumieniowego

**Input** :  $D$ : Training dataset,  
 $n$ : Size of the mini-batch,  
 $T$ : Total number of iterations

- 1 Construct a statistical population  $P_D$  from the training dataset  $D$ ;
- 2 Initialize the probabilities  $p_i$  for each instance  $i$  in  $P_D$  according to the chosen method;
- 3 **for**  $\tau = 1, \dots, T$  **do**
- 4     Sample a new mini-batch  $B_\tau$  of size  $n$  from the population  $P_D$ ;
- 5     Train the network on the mini-batch  $B_\tau$ ;
- 6     Update the probabilities  $p_i$  for each instance in  $B_\tau$  using the chosen method;
- 7     Normalize the probabilities  $p_i$  of the population  $P_D$  to ensure they sum up to 1;
- 8 **end**

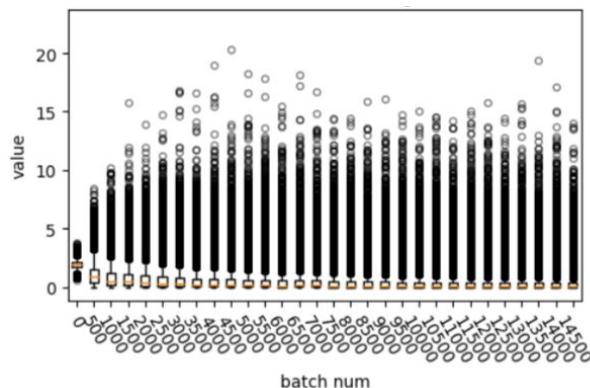
Method name	Initialization values	Update method
Only Wrongly Classified (OWC)	$p_i \leftarrow \frac{1}{N}$	Probabilities of misclassified examples are increased by a predetermined value $\alpha$
Loss Based (LB)	$p_i \leftarrow \mathcal{L}(f(\mathbf{x}_i), d_i)$	$p_i \leftarrow \mathcal{L}(f(\mathbf{x}_i), d_i)$
Normalized Loss Based (NLB)	$p_i \leftarrow \tanh \mathcal{L}(f(\mathbf{x}_i), d_i)$	$p_i \leftarrow \tanh \mathcal{L}(f(\mathbf{x}_i), d_i)$

# Chaotyczne zachowanie rozkładu prawdopodobieństwa

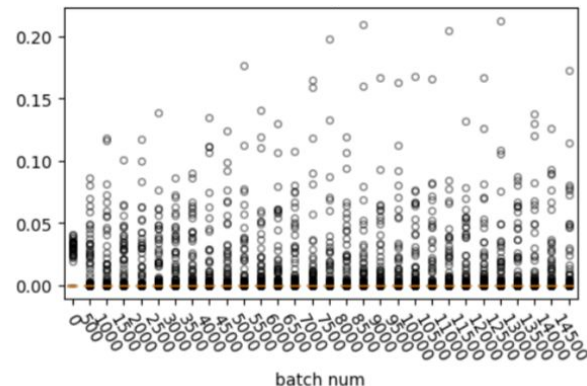
## Problem:

Normalizacja wartości PoD powoduje nieprzewidywalne fluktuacje ich wartości, przez co rozkład prawdopodobieństwa nie przybliża kształtem faktycznego rozkładu trudności elementów.

True distribution



Stored PoD values



# Mechanizm wartości bazowych

**Problem:** Normalizacja wprowadza rozbieżność kształtu rozkładów

**Pomysł:** Zachować wzajemną porównywalność trudności elementów

$$\mathcal{T}_D = \{(x_i, d_i, b_i)\} \quad b_i \in \mathbb{R}_+$$

Wartości bazowe (ang. *Base-Values*) odzwierciedlają trudność elementów uczących: im trudniejszy przykład, tym wyższa wartość bazowa.

Wprowadzają relację częściowego porządku nad zbiorem uczącym, tworząc wiarygodny mechanizm porównania względnej istotności elementów uczących.

Rozkład prawdopodobieństwa jest wprost proporcjonalny do wartości bazowych

$$p_i = \frac{b_i}{\sum_{j=1}^N b_j}$$

Konstrukcja wartości bazowych skutkuje bardziej stabilnym rozkładem prawdopodobieństwa i wiarygodniejszym sposobem porównania względnej istotności elementów uczących.

# Nowe metody losowania: PLB and PEB

## Persistent Loss Based Approach (PLB)

- Funkcja kosztu jako metryka niepewności modelu
- Brak normalizacji wartości

## Persistent Entropy Based Approach (PEB)

- Entropia przewidzianego przez model wektora przynależności do klas jako metryka niepewności tego modelu
- Nadaje się tylko do klasyfikacji
- Brak normalizacji wartości

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad H(P) = - \sum_{k=1}^K P_k \log(P_k)$$

# Algorytm PLB / PEB

**Input** :  $D$ : Training dataset of size  $N$ ,  
 $n$ : Size of the mini-batch,  
 $T$ : Total number of iterations,  
 $L$ : The loss function

1 Construct a set of triplets  $\mathcal{T}_D$  from  $D$

// Initializing Base-Values in  $\mathcal{T}_D$

2 **foreach**  $i = 1, 2, \dots, N$  **do**

3     Initialize  $b_i$  according to formula (5) or (6)

4 **end**

5 **for**  $\tau = 1, 2, \dots, T$  **do**

6     Sample a new mini-batch  $B_\tau$  from  $\mathcal{T}_D$  according to the  
   probability mass function derived from formula (2)

7     Train the network on the mini-batch  $B_\tau$

// Updating Base-Values

8     **foreach** instance  $s_j$  in  $B_\tau$  **do**

9         Update  $b_{i_j}$  according to formula (5) or (6)

10     **end**

11 **end**

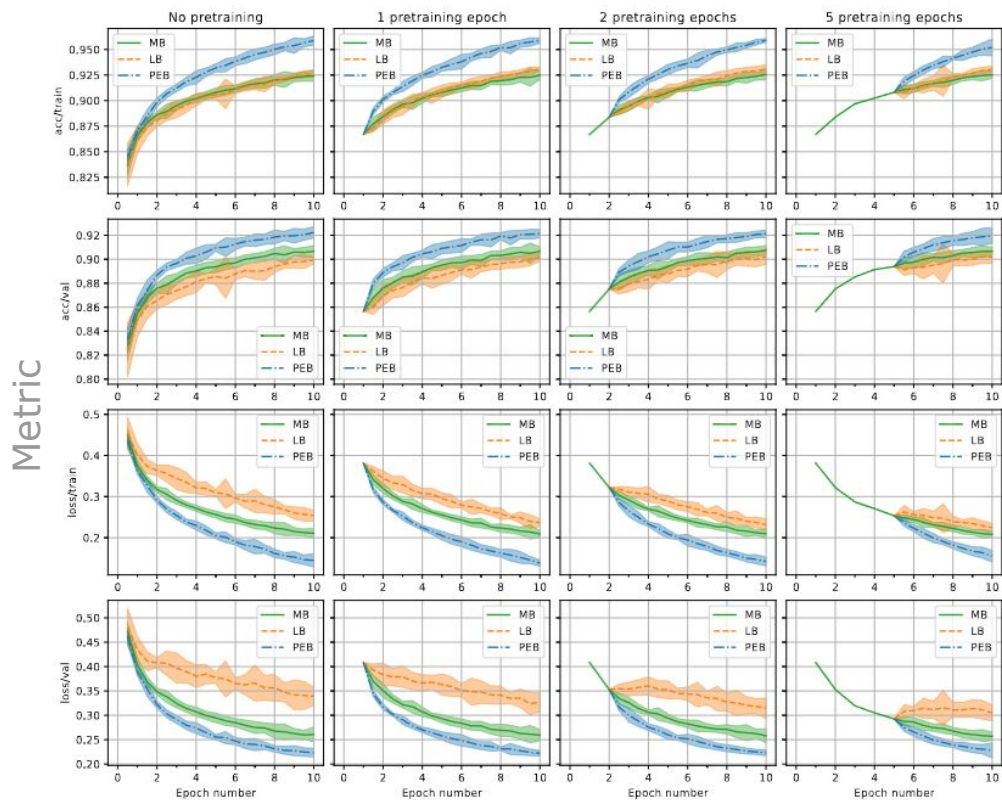
$$b_i = L(x_i, d_i) \quad (5)$$

$$b_i = H(\sigma(y_i)) \quad (6)$$

$$p_i = \frac{b_i}{\sum_{j=1}^N b_j} \quad (2)$$

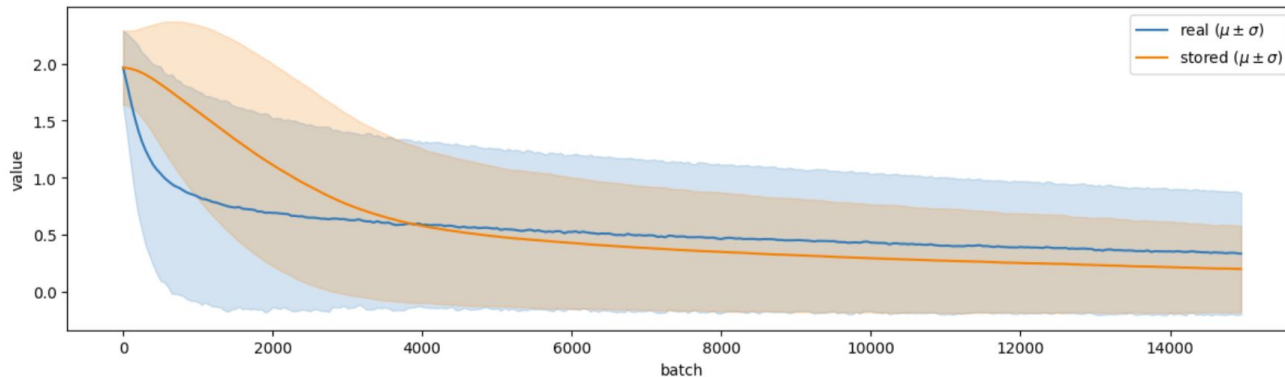
# Wyniki eksperymentów

## Experiment

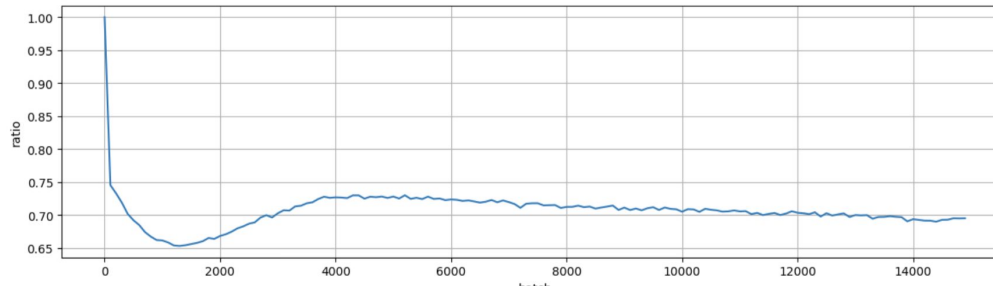


# Czy zachowana jest względna porównywalność trudności elementów?

Real and stored basevalues distribution shift comparison in PLB training



Estimated percentage of correctly ordered stored basevalues compared to real basevalues



# Future Research Directions

## Narzut obliczeniowy

Obliczanie wartości bazowych i losowanie kolejnych mini-batchy wymaga dodatkowych obliczeń, co ma wpływ na czas uczenia.

## Nieaktualne wartości bazowe

Wartości bazowe są aktualne wyłącznie dla elementów uczących z ostatnio stworzonego mini-batcha.

## Robust Training

Badanie wydajności podejścia strumieniowego w zagadnieniach Robust Training (np. uczeniu w obecności ataków adversarialnych lub zaszumionych etykiet).

## Nowe techniki próbkowania

Opracowywanie nowych technik próbkowania w celu dalszej poprawy wydajności metody w różnych scenariuszach.



**CEAI**  
Center of Excellence in Artificial Intelligence

artic

# Dziękuję za uwagę

---



Mateusz Wojtulewicz



wojtulewicz@agh.edu.pl