

В. В. Півошенко
Ю. С. Здітовецький
С. Г. Кривогубченко
Ю. Ю. Іванов

ОПТИМІЗАЦІЯ ГЛИБОКОЇ НЕЙРОННОЇ МЕРЕЖІ НА ОСНОВІ ВИКОРИСТАННЯ ЦИКЛІЧНО-КОСИНУСОЇДНОЇ ЗМІНИ КОЕФІЦІЄНТА ШВИДКОСТІ НАВЧАННЯ ТА АНСАМБЛЮВАННЯ МОДЕЛЕЙ

Вінницький національний технічний університет

Анотація

У даній роботі проаналізовано особливості використання циклічно-косинусоїдної зміни коефіцієнта швидкості навчання для створення ансамблю комбінаційних моделей автокодера та згорткової нейронної мережі.

Ключові слова: машинне навчання, коефіцієнт швидкості навчання, ансамбль моделей, автокодер, згорткова нейронна мережа.

Abstract

In this work have been analyzed the features of using the cosine annealing learning rate for creating an ensemble for combination model of autoencoder and convolutional neural network.

Keywords: machine learning, learning rate, ensemble of models, autoencoder, convolutional neural network.

Вступ

Останніми роками спостерігається вибух інтересу до нейронних мереж, які успішно застосовуються в різних галузях науки і техніки. Нейронна мережа навчається за допомогою певного алгоритму, у ході роботи якого вибірка даних використовується для коригування ваг мережі таким чином, щоб мінімізувати помилку прогнозу на навчальній множині [1]. Стохастичний градієнтний спуск (СГС) [2] та його модифікації [3] стали класичними підходами оптимізації глибоких нейронних мереж (ГНМ). Популярність СГС можна пояснити здатністю уникати потрапляння до «сідлових» точок та локальних мінімумів, прямуючи до глобального екстремуму. Проте, встановлено [4], що кількість можливих локальних мінімумів зростає експоненціально зі збільшенням параметрів, кількість яких у сучасних ГНМ може бути величезною. Це явище сприяє тому, що СГС не досягає глобального екстремуму, а потрапляє у певний локальний мінімум. Тому дві моделі з однаковою архітектурою, оптимізовані різними ініціалізаціями СГС, сходяться до різних локальних оптимумів. Такий підхід дозволяє створити ансамблі, в яких навчаються декілька моделей з визначеною архітектурою, але з різною ініціалізацією СГС [4, 5] та усередненням виходів, що призводить до зниження частоти помилок. Незважаючи на свої очевидні переваги, використання ансамблю ГНМ є досить ресурсомістким.

Для вирішення даної проблеми було запропоновано низку підходів [5-7], але їх ефективність досліджувалася лише на ГНМ із класичною та простою архітектурою, такою як MobileNet, MobileNetV2. Це не може підтверджувати ефективність даних підходів на ГНМ із складною будовою та великою кількістю параметрів, наприклад комбінаційна модель автокодера та згорткової нейронної мережі.

Тому *метою* роботи є дослідження особливостей використання циклічно-косинусоїдної зміни коефіцієнта швидкості навчання (КШН) мережі для створення ансамблю комбінаційних моделей автокодера та згорткової нейронної мережі у ході розв'язання задачі детектування пневмотораксу.

Результати дослідження

Застосовуємо СГС M разів для пошуку локальних мінімумів, зберігаючи ваги моделі кожного разу після того, як процес конвергується (рисунок 1). Щоб уникнути поточного локального мінімуму, з кожним перезапуском виконується зміна КШН.

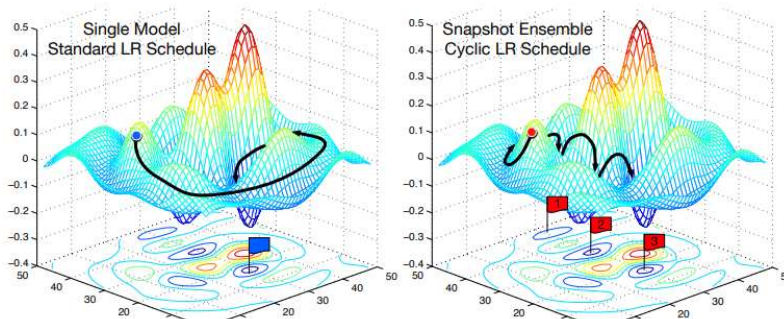


Рисунок 1 – Зліва: типовий процес оптимізації СГС; справа: використання локальних мінімумів для ансамблювання моделі

Для руху до локальних мінімумів використано циклічну зміну КШН [7], який зменшується під час руху до першого локального екстремуму, а потім різко збільшується, що і виводить модель з локального оптимуму. Даний процес повторюється багаторазово. Формально КШН α можна обчислити за виразом

$$\alpha(t) = f(\text{mod}(t - 1, \lceil T/M \rceil)), \quad (1)$$

де t – номер епохи; f – монотонно спадна функція; T – загальна кількість епох; M – кількість запусків.

Отже, процес навчання поділено на M циклів, кожен з яких починається з великого значення КШН, яке поступово зменшується. Велике значення КШН $\alpha = f(0)$ дозволяє моделі, уникнути «сідлових» точок, а зменшення КШН $\alpha = f(\lceil T/M \rceil)$ забезпечує рух до локального мінімуму. У ході наших експериментів використано циклічно-косинусоїдну зміну КШН, графік якої наведено на рисунку 2. Вираз для розрахунку задається у формі (α_0 – початкове значення КШН)

$$\alpha(t) = \frac{\alpha_0}{2} \left(\cos\left(\frac{\pi \text{mod}(t-1, \lceil \frac{T}{M} \rceil)}{\lceil \frac{T}{M} \rceil}\right) + 1 \right). \quad (2)$$

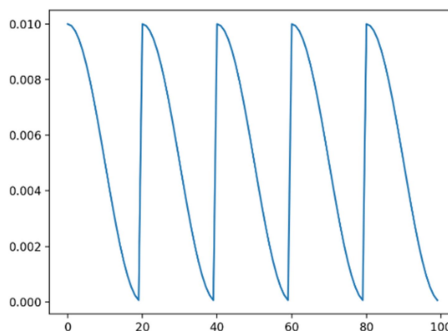


Рисунок 2 – Графік циклічно-косинусоїдної зміни КШН

Впродовж навчання дана функція змінює значення КШН від α_0 до $f(\lceil T/M \rceil) \approx 0$. Після навчання можна отримати набір ваг $W = [W_0, W_1, \dots, W_M]$, які перед процесом прогнозування усереднюються та завантажуються у модель.

У даній роботі для розв’язання задачі детектування пневмотораксу запропоновано ансамблювання з використанням комбінаційної моделі автокодера та згорткової нейронної мережі та вибрано два критерії для оцінювання її ефективності: час навчання і валідаційний коефіцієнт Жаккара J (таблиця 1).

Таблиця 1 – Результати дослідження

Тип ансамблювання	Час, сек	J , %
Класичне ансамблювання, 3 моделі	91500	83 %
Ансамблювання з використанням циклічно-косинусоїдної зміни КШН та усереднення вагів	29790	89 %

Висновки

У даній роботі проаналізовано використання циклічно-косинусоїдної зміни КШН для створення ансамблю комбінаційних моделей автокодера та згорткової нейронної мережі. Визначено, що запропонована модель є ефективнішою за класичне ансамблювання у ході розв’язання практичної задачі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Low Computational Complexity Algorithm for Recognition Highly Corrupted QR Codes Based on Hamming-Lippmann Neural Network / R.N. Kvyetnyy, Yu.Yu. Ivanov, V.V. Pivoshenko and others // Przegląd Elektrotechniczny. – Poland, 2019. – P. 162-166.
2. Bottou L. Large-Scale Machine Learning with Stochastic Gradient Descent / L. Bottou // Proceedings of COMPSTAT 2010. – 2010. – P. 177-186.
3. Rozložník M. Solution Approaches for Saddle-Point Problems / M. Rozložník // Nečas Center Series Saddle-Point Problems and Their Iterative Solution. – 2018. – P. 33-39.
4. Kawaguchi K. Effect of Depth and Width on Local Minima in Deep Learning / K. Kawaguchi, J. Huang, L.P. Kaelbling // Neural Computation. – 2019. – P. 1462-1498.
5. Caruana R. Getting the Most Out of Ensemble Selection / R. Caruana, A. Munson, A. Niculescu-Mizil // Sixth International Conference on Data Mining (ICDM'06). – 2006. – 12 p.
6. Zagoruyko S. Wide Residual Networks / S. Zagoruyko, N. Komodakis // Proceedings of the British Machine Vision Conference 2016. – 2016. – 15 p.
7. Loshchilov I. Comparison-Based Optimizers Need Comparison-Based Surrogates / I. Loshchilov, M. Schoenauer, M. Sebag // Parallel Problem Solving from Nature, PPSN XI. – 2010. – P. 364-373.

Півошенко Володимир Володимирович — студент групи ІСі-166, факультет комп’ютерних систем і автоматики, Вінницький національний технічний університет, м. Вінниця, e-mail: volodymyr.pivoshenko@gmail.com.

Здітовецький Юрій Сергійович — студент групи ІСТ-19м, факультет комп’ютерних систем і автоматики, Вінницький національний технічний університет, м. Вінниця.

Кривогузченко Сергій Григорович — канд. техн. наук, доцент кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця.

Іванов Юрій Юрійович — канд. техн. наук, доцент кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, e-mail: Yura881990@i.ua.

Pivoshenko Volodymyr V. — student, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, e-mail: volodymyr.pivoshenko@gmail.com.

Zditovetskiy Yuriy S. — student, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, e-mail: volodymyr.pivoshenko@gmail.com.

Krivogubchenko Sergiy G. — Cand. Sc. (Eng), Senior Lecturer, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia.

Ivanov Yuriy Yu. — Cand. Sc. (Eng), Senior Lecturer, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: Yura881990@i.ua.