

# ОСОБЛИВОСТІ ЗАСТОСУВАННЯ МЕХАНІЗМУ УВАГИ (ATTENTION) У КЛАСИФІКАЦІЇ МУЗИЧНИХ ТВОРІВ

Вінницькій національний технічний університет

## Анотація

У статті розглянуто механізм уваги (Attention) як інноваційний підхід до класифікації музичних творів. Описано процес визначення релевантних частин аудіосигналу для покращення точності моделі. Розглянуто функції оцінки важливості, нормалізації ваг та обчислення контекстного вектора. Наведено приклад використання уваги в рекурентних нейронних мережах (RNN) та її переваги для аналізу складних музичних структур.

**Ключові слова:** класифікація музики, механізм уваги, машинне навчання, аудіосигнал.

## Abstract

The article examines the attention mechanism as an innovative approach to music classification. It describes the process of identifying relevant parts of audio signals to improve model accuracy. Importance scoring, attention weight normalization, and context vector computation are discussed. An example of attention implementation in recurrent neural networks (RNN) is provided, showcasing its advantages for analyzing complex musical structures.

**Keywords:** music classification, attention mechanism, machine learning, audio signal.

Механізм уваги (Attention) [1] став справжнім проривом у сфері машинного навчання, відкриваючи нові горизонти в обробці послідовних даних. У контексті класифікації музичних творів цей підхід дозволяє моделям зосереджуватися на найбільш важливих частинах аудіосигналу, що мають вирішальне значення для правильного визначення жанру або стилю композиції.

Уявімо, що ми працюємо з послідовністю векторів ознак  $X = [x_1, x_2, \dots, x_T]$ , де  $T$  – довжина аудіозапису. Кожен вектор  $x_t$  представляє набір характеристик музичного сигналу в момент часу  $t$ . Завдання моделі полягає в тому, щоб визначити, які з цих моментів є найбільш релевантними для класифікації.

Для цього вводиться функція оцінки важливості [2], яка для кожного елемента  $x_t$  обчислює скалярне значення  $e_t$ :

$$e_t = \text{score}(x_t, s) \quad (2.7)$$

де:

$s$  — контекстний вектор або прихований стан моделі;

$\text{score}$  — функція, яка може бути реалізована різними способами, наприклад, як скалярний добуток або невелика нейронна мережа.

Ці оцінки відображають, наскільки важливим є кожен момент часу для поточного завдання.

Потім ці оцінки нормалізуються за допомогою softmax-функції, щоб отримати ваги уваги  $\alpha_t$ :

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^T \exp(e_k)} \quad (2.8)$$

Ці ваги мають властивість сумуватися до 1, що інтерпретується як ймовірність або ступінь уваги, яку модель приділяє кожному елементу послідовності. Після цього обчислюється контекстний вектор  $c$ , який є зваженою сумою вхідних ознак:

$$c = \sum_{t=1}^T \alpha_t x_t \quad (2.9)$$

Цей контекстний вектор містить агреговану інформацію про найбільш значущі частини аудіосигналу і використовується для подальшої класифікації. Наприклад, він може бути поданий на вхід повноз'язаного шару з функцією softmax для передбачення ймовірностей належності до різних жанрів.

Функція втрат для навчання моделі зазвичай вибирається як крос-ентропія:

$$L = -\sum_{t=1}^T \sum_{k=1}^K y_{tk} \log(\hat{y}_{tk}) \quad (2.10)$$

де:

$N$  — кількість зразків у навчальному наборі;

$K$  — кількість класів (жанрів);

$y_{ik}$  — істинна мітка класу;

$\hat{y}_{tk}$  — прогнозована ймовірність.

Використання механізму уваги має кілька важливих переваг. По-перше, модель стає здатною адаптивно визначати, які частини аудіосигналу є найбільш релевантними для класифікації. Це особливо цінно для музичних творів зі складною структурою, де важливі елементи можуть бути розподілені нерівномірно по всій тривалості треку. По-друге, механізм уваги сприяє поліпшенню точності моделі, оскільки зменшується вплив нерелевантної або зашумленої інформації. По-третє, ваги уваги  $a_t$  надають можливість інтерпретувати роботу моделі, оскільки вони вказують на те, які саме фрагменти аудіо найбільше вплинули на кінцеве рішення.

Розглянемо приклад використання механізму уваги в моделі для класифікації музики. Нехай ми маємо модель на основі рекурентної нейронної мережі (RNN), яка обробляє послідовність ознак  $x_t$  і генерує приховані стани  $h_t$ :

$$h_t = RNNCell(x_t, h_{t-1}) \quad (2.11)$$

Функція оцінки важливості може бути реалізована як адитивна увага:

$$e_t = v \tanh(W h_t + b) \quad (2.12)$$

де  $W, v, b$  — параметри, що навчаються. Після обчислення ваг уваги  $a_i$  і контекстного вектора  $c$ , модель робить передбачення:

$$\hat{y} = \text{softmax}(W_c c + b_c) \quad (2.13)$$

де  $W_c, b_c$  — параметри вихідного шару.

Одним із викликів при використанні механізму уваги є підвищена обчислювальна складність, особливо для довгих послідовностей. Це може вимагати оптимізації коду та використання потужних апаратних ресурсів. Однак, завдяки сучасним технологіям та графічним процесорам (GPU), ці проблеми можна ефективно вирішувати.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Могілянський математичний журнал. Електронний ресурс. Режим доступу: <http://mmj.ukma.edu.ua/article/view/275054>.
2. Пахомов К. О. Використання механізмів уваги для аналізу аудіосигналів. Вісник Київського національного університету імені Тараса Шевченка. Серія: Інформатика. — 2021. — № 2. — С. 45–51.

**Каспров Владислав Олександрович** – студент групи 2ПІ-23м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, [vlad.kasprov@gmail.com](mailto:vlad.kasprov@gmail.com).

**Kasprov Vladyslav Oleksandrovyich** – a student of group 2PI-23m, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, [vlad.kasprov@gmail.com](mailto:vlad.kasprov@gmail.com).