 МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

УНИВЕРСИТЕТ

Институт №8

«Компьютерные науки и прикладная математика» Кафедра №806

«Вычислительная математика и программирование»

Отчет по лабораторной работe

MPI + PyTorch

по учебной дисциплине

«Параллельные и распределенные вычисления»

Выполнил:

Студент 1-го курса

Гр. М80-114M-22

Гордеев Н. М.

.

(подпись, дата) .

Москва 2022

Оглавление

[1. Постановка задачи 2](#_Toc122522811)

[2. Теория 2](#_Toc122522812)

[3. Метод решения 2](#_Toc122522813)

[4. Результат работы программы 3](#_Toc122522814)

[5. Вывод: 3](#_Toc122522815)

# 1. Постановка задачи

Распараллелить с помощью mpi4py задачу обучения нейронной сети в PyTorch. Получить таким образом n моделей и ансамблем дать предсказание для тестового набора данных.

# 2. Теория

Message Passing Interface (MPI, интерфейс передачи сообщений) — программный интерфейс (API) для передачи информации, который позволяет обмениваться сообщениями между процессами, выполняющими одну задачу.

Ансамбль алгоритмов (методов) — метод, который использует несколько обучающих алгоритмов с целью получения лучшей эффективности прогнозирования, чем можно было бы получить от каждого обучающего алгоритма по отдельности.

PyTorch — фреймворк машинного обучения для языка Python с открытым исходным кодом, созданный на базе Torch. Используется для решения различных задач: компьютерное зрение, обработка естественного языка.

# 3. Метод решения

Будет запущено 4 процесса с помощью mpi4py. На 0 процессоре набор данных Diabetic\_Retinopathy будет разбит на 9 частей (4 train, 4 wall, 1 test) и информация об разбиении будет записана в файл. Далее все процессы будут учится 100 эпох на своих train и wall частях. После обучения все модели дадут предсказание для test и отошлют их на 0 процесс. На 0 процессе ответы 4 моделей преобразуется в 1 ответ ансамбля. Будет посчитана метрика acc ответов на тестовый набор данных каждой модели и ансамбля.

# 4. Результат работы программы

Записан в файле log.txt

Выжимка:  
Rank 0 Acc: 0.7337

Rank 1 Acc: 0.7224

Rank 2 Acc: 0.7321

Rank 3 Acc: 0.7300

General Acc: 0.7376

# 5. Вывод:

Метод ансамбля 4 независимых моделей дал крошечный прирост к качеству, но если этот набор данных разбить не на 9 равных частей а на 2 с соотношением 80:20 и обучить 1 модель на большей части, то можно добиться гораздо более высокого качества работы.