

УДК 004.942

Момотюк В.–ст. гр. СП-42

Тернопільський національний технічний університет імені Івана Пулюя

РОЗРОБКА ПРОГРАМНОЇ ПЛАТФОРМИ ДЛЯ МОНІТОРИНГУ РИЗИКІВ СТОХАСТИЧНИХ СИСТЕМ З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖЕВИХ МОДУЛІВ

Науковий керівник: к.т.н., доцент Стоянов Ю. М.

Momotiuk V.

Ternopil Ivan Puluji National Technical University

DEVELOPMENT OF A SOFTWARE PLATFORM FOR RISK MONITORING OF STOCHASTIC SYSTEMS USING NEURAL NETWORK MODULES

Supervisor: Ph.D. in Engineering, associate professor Stoianov Ju .M.

Ключові слова: програмна платформа, моніторинг ризиків, стохастичні системи, штучні нейронні мережі, стохастичні обчислення, глибоке навчання

Keywords: software platform, risk monitoring, stochastic systems, artificial neural networks, stochastic computing, deep learning

Сьогодні штучні нейронні мережі (ШНМ) та алгоритми глибокого навчання є основою для вирішення складних завдань класифікації, розпізнавання образів та аналітики в системах штучного інтелекту. Проте розмір сучасних глибоких нейронних мереж часто перевищує тисячі вузлів, що робить суто програмні рішення на стандартних процесорах надто повільними для застосунків, які потребують тривалого навчання та миттєвого використання (інференсу). Серйозним викликом у застосунках реального часу стає апаратна реалізація багатопаралельних нейронних мереж (наприклад, багатопаралельних перцептронів).

Традиційні архітектури вимагають використання десятків тисяч складних апаратних множників, що призводить до величезного споживання енергії та надмірного використання ресурсів кристала, тому не підходять для портативних рішень. Серед різноманітних методів оптимізації величезну увагу дослідників привернули обчислювальні архітектури на базі стохастичних обчислень (Stochastic Computing, SC).

Цей підхід дозволяє замінити складні арифметичні операції на прості логічні вентиля (наприклад, вентиля XNOR замість великих мультиплікаторів), що кардинально знижує апаратні витрати та робить систему енергоефективною. Однак, функціонування систем на базі стохастичних обчислень містить у собі серйозні внутрішні ризики та обмеження.

Головною проблемою є повільна збіжність: для генерації відносно точного результату системі та традиційним оцінювачам ймовірності (PE) потрібні мільйони тактових імпульсів. Крім того, виникають проблеми "масштабування вниз" (scale-down) та висока ймовірність флуктуаційних похибок при математичних операціях. Ідеальною платформою для розгортання таких нейромереж є програмовані логічні інтегральні

схеми (ПЛІС, або FPGA) завдяки їхній реконфігурованості та можливості швидких паралельних обчислень.

Проте обмежена кількість пінів вводу-виводу на ПЛІС створює вузьке місце під час передачі масивів даних (наприклад, зображень). Щоб нівелювати обчислювальні ризики (затримки, втрату точності), інженерам необхідно впроваджувати складні конвеєрні архітектури та послідовний ввід даних. У зв'язку з цим, перед дорогим розгортанням таких стохастичних систем на реальних апаратних пристроях, виникає гостра потреба у програмних інструментах їх тестування та дата-орієнтованої верифікації.

Актуальність теми полягає у необхідності створення ефективної програмної платформи, яка дозволить моделювати роботу нейромережових модулів на базі стохастичних обчислень, моніторити пов'язані з ними обчислювальні ризики (флуктуаційні похибки, затримки збіжності) та оптимізувати архітектурні рішення до етапу їх фізичного впровадження у вбудованих пристроях.

Метою роботи є розробка програмної платформи для моніторингу ризиків стохастичних систем з використанням нейромережових модулів, що дозволяє моделювати стохастичні обчислювальні процеси, оцінювати рівень їх точності та забезпечувати надійність алгоритмів перед їх застосуванням у системах реального часу.

Об'єктом дослідження є процеси функціонування та оцінки надійності стохастичних динамічних систем на базі нейромережових модулів в умовах складних обчислювальних обмежень.

Предметом дослідження є методи, алгоритми та програмні засоби моніторингу ризиків (флуктуаційних похибок, часу збіжності) у штучних нейронних мережах, побудованих на принципах стохастичних обчислень. Для досягнення поставленої мети в роботі сформульовано та вирішено такі основні завдання:

- Провести аналіз існуючих методів реалізації глибоких нейронних мереж (ГНМ) та виявити ключові ризики і проблеми використання стохастичних обчислень (SC) в системах реального часу.
- Розробити математичні та алгоритмічні моделі нейромережових модулів, які функціонують на принципах стохастичних бітових потоків, з урахуванням новітніх методів прискорення збіжності.
- Спроекувати архітектуру програмної платформи, здатну симулювати послідовний ввід даних та конвеєрну обробку, характерну для апаратних обмежень ПЛІС.
- Здійснити практичну розробку програмного рішення для дата-орієнтованої верифікації та моніторингу обчислювальних ризиків.
- Провести тестування розробленої платформи на модельних наборах даних (зокрема для задач обробки зображень) та порівняльний аналіз із сучасними (state-of-the-art) аналогами.

Практичне значення отриманих результатів. В роботі розроблено програмну платформу, яка дозволяє комплексно моніторити ризики стохастичних нейромереж. Її практичне значення підтверджується тим, що платформа дозволяє верифікувати та тестувати архітектурні рішення, які на етапі апаратної реалізації здатні забезпечити скорочення використання ресурсів більш ніж на 82% та мінімальне енергоспоживання. Крім того, завдяки точному моніторингу та моделюванню синхронізації, розроблений інструментарій допомагає знизити середній рівень помилки змодельованої нейронної мережі на 2%, що робить систему ідеальною для підготовки надійних Edge-AI рішень для обробки зображень.