Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0825U000334

Особливі позначки: відкрита

Дата реєстрації: 22-01-2025

Статус: Запланована

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Іванюк Андрій Олегович
2. Andrii Ivaniuk

Кваліфікація: д.філософ, 113

Ідентифікатор ORCID ID: 0000-0002-4189-3787

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 113

Назва наукової спеціальності: Прикладна математика

Галузь / галузі знань: математика та статистика

Освітньо-наукова програма зі спеціальності: Прикладна математика

Дата захисту: 13-12-2024

Спеціальність за освітою: Інженерія програмного забезпечення

Місце роботи здобувача:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 7057

Повне найменування юридичної особи: Національний університет "Києво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, буд. 2, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний університет "Києво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, буд. 2, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 27, 27.03.17, 27.03.21

Тема дисертації:

1. Дослідження взаємозв'язків у даних з використанням шту- чних нейронних мереж
2. Study of relationships in data using artificial neural networks

Реферат:

1. Іванюк А.О. Дослідження взаємозв'язків у даних з використанням штучних нейронних мереж — Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії у галузі знань 11 "Математика та статистика" за спеціальністю 113 "Прикладна математика". — Національний університет «Києво-Могилянська академія», Київ, 2024. Ця дисертація зосереджена на вивченні зв'язків у даних за допомогою застосування штучних нейронних мереж. Ці зв'язки можуть бути представлені в різних формах, і моделюватись по-різному. Їх правильне моделювання є ключовим для успішного вирішення різноманітних завдань, таких як класифікація, регресія та генеративне моделювання. У сучасних нейронних мережах широко використовуються стандартні метрики для оцінки їх продуктивності, наприклад, класифікаційна точність, середньоквадратична похибка тощо. Проте, високі показники цих метрик не гарантують відсутності помилок або вразливостей у моделях. Моделі можуть видавати помилкові результати з високим рівнем впевненості, особливо при взаємодії з адверсаріальними прикладами — спеціально створеними вхідними даними, які вводять модель в оману. Це дослідження стосується цієї

важливої проблеми шляхом детального вивчення кількісної оцінки невизначеності та стійкості нейронних мереж до адверсаріальних атак. Використовуючи адверсаріальні дані як інструмент, ця робота спрямована на поглиблення розуміння надійності моделей та розроблення більш стійких систем на основі нейронних мереж, які можуть протистояти різноманітним атакам та забезпечувати стабільну продуктивність у реальних застосуваннях. Досліджуючи адверсаріальні взаємозв'язки та патерни в даних, ця робота має на меті використовувати їх як метрику генералізації для виявлення слабких місць моделей та оцінки їх здатності до узагальнення. Розуміння того, як моделі реагують на суперечливі збурення, відкриває унікальний погляд на їх внутрішню структуру та механізми прийняття рішень. Це дозволяє не лише виявляти вразливі місця, але й розробляти методи для їх усунення, що підвищує загальну надійність та ефективність моделей. У рамках цього дослідження вивчаються різні параметризації нейронних мереж для моделювання послідовностей та їх вплив на продуктивність моделей і стійкість до адверсаріальних атак. Особлива увага приділяється новим архітектурам та активаційним функціям, які можуть покращити здатність моделей до генералізації та їхню стійкість. Адверсаріальна стійкість розглядається як важлива метрика для виявлення слабких місць моделей та оцінки їх загальної ефективності. Дослідження охоплює ефективні параметризації для різних типів вхідних даних, включаючи зображення, мовні сигнали та текст. Застосовуються ці параметризації до різних завдань машинного навчання, таких як класифікація зображень, моделювання мови та регресія на основі латентних дифузійних моделей. Проведені експерименти спрямовані на виявлення того, як різні стратегії параметризації можуть покращити продуктивність моделей, зберігаючи або навіть підвищуючи їх стійкість до адверсаріальних атак. Отримані результати надають важливі знання для розробки більш надійних та здатних до генералізації моделей машинного навчання. Це сприяє прогресу у цій галузі шляхом виявлення оптимальних технік параметризації, які збалансовують продуктивність та стійкість, та можуть бути застосовані у широкому спектрі практичних задач. Загалом, ця дисертація робить вагомий внесок у розуміння та покращення стійкості нейронних мереж до адверсаріальних атак, пропонуючи нові підходи до параметризації та моделювання, які можуть бути застосовані у різних сферах машинного навчання. Результати цього дослідження можуть стати основою для розробки більш надійних та ефективних моделей, здатних забезпечувати високу продуктивність та безпеку у реальних застосуваннях. Проведені експерименти підтверджують, що використання розглянутих параметризацій може підвищити точність класифікації, але також виявляють складності, пов'язані з їх адверсаріальним тренуванням. Подальші дослідження у цьому напрямку можуть призвести до створення моделей, які не лише демонструють високу продуктивність, але й є стійкими до різноманітних атак, що є критично важливим у сучасному світі, де безпека та надійність моделей машинного навчання набувають все більшого значення. Ключові слова: адверсаріальна стійкість, адверсаріальні приклади, адверсаріальне очищення, механізм уваги, параметризація моделей, дифузійне моделювання, автокодувальники, обробка сигналів, адаптивні алгоритми, алгоритми оптимізації, функції активації, регуляризація нейронні мережі, штучна нейронна мережа, алгоритм, згорткова нейронна мережа, параметри, похибка машинне навчання, класифікація, регресія, генеративне моделювання, комп'ютерний зір, обробка природної мови, аудіо моделювання.

2. Ivaniuk A.O. Study of relationships in data using artificial neural networks. — Qualified research work (manuscript). Dissertation to obtain the scientific degree of Doctor of Philosophy in the Field of Study 11 "Mathematics and statistics", Programme Subject Area 113 "Applied mathematics". — National University of Kyiv-Mohyla Academy, Kyiv, 2024. This dissertation focuses on studying relationships in data through the application of artificial neural networks. These relationships can be represented in various forms and modeled in different ways. Correct modeling of these relationships is key to successfully solving a variety of tasks, such as classification, regression, and generative modeling. In modern neural networks, standard metrics are widely used to evaluate their performance, such as classification accuracy, mean squared error, and so on. However, good values of these metrics do not guarantee the absence of errors or vulnerabilities in models. Models can produce erroneous results with a high level of confidence, especially when interacting with adversarial examples—specially crafted input data that mislead the model. This research addresses this important problem by conducting a detailed study of quantitative assessment of uncertainty and the robustness of neural networks to adversarial attacks. By using

adversarial data as a tool, this work aims to deepen the understanding of model reliability and to develop more robust neural networkbased systems that can withstand various attacks and provide stable performance in real-world applications. By investigating adversarial relationships and patterns in data, this work aims to use them as a metric of generalization to identify model weaknesses and assess their ability to generalize. Understanding how models respond to conflicting perturbations offers a unique perspective on their internal structure and decisionmaking mechanisms. This allows not only for the identification of vulnerabilities but also for the development of methods to eliminate them, thereby enhancing the overall reliability and efficiency of models. As part of this research, various parameterizations of neural networks for sequence modeling are studied, as well as their impact on model performance and robustness to adversarial attacks. Special attention is paid to new architectures and activation functions that can improve models' ability to generalize and their robustness. Adversarial robustness is considered an important metric for identifying model weaknesses and evaluating their overall effectiveness. The research encompasses effective parameterizations for different types of input data, including images, speech signals, and text. These parameterizations are applied to various machine learning tasks, such as image classification, language modeling, and regression based on latent diffusion models. The experiments conducted aim to identify how different parameterization strategies can improve model performance while maintaining or even enhancing their robustness to adversarial attacks. The results obtained provide important insights for developing more reliable and generalizable machine learning models. This advances the field by identifying optimal parameterization techniques that balance performance and robustness and can be applied in a wide range of practical tasks. Overall, this dissertation makes a significant contribution to understanding and improving the robustness of neural networks to adversarial attacks by proposing new approaches to parameterization and modeling that can be applied in various fields of machine learning. The results of this research can serve as a foundation for developing more reliable and efficient models capable of ensuring high performance and security in real-world applications. The experiments conducted confirm that using the considered parameterizations can enhance classification accuracy but also reveal complexities associated with their adversarial training. Further research in this direction may lead to the creation of models that not only demonstrate high performance but are also robust to various attacks, which is critically important in today's world where the security and reliability of machine learning models are becoming increasingly significant. Keywords: adversarial robustness, adversarial examples, adversarial purification, attention mechanism, model parameterization, diffusion modeling, auto-encoders, signal processing, adaptive algorithms, optimization algorithms, activation functions, regularization, neural networks, artificial neural network, algorithm, convolutional neural network, parameters, error, machine learning, classification, regression, generative modeling, computer vision, natural language processing, audio modeling.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Фундаментальні наукові дослідження з найбільш важливих проблем розвитку науково-технічного, соціально-економічного, суспільно-політичного, людського потенціалу для забезпечення конкурентоспроможності України у світі та сталого розвитку суспільства і держави

Стратегічний пріоритетний напрям інноваційної діяльності: Не застосовується

Підсумки дослідження: Теоретичне узагальнення і вирішення важливої наукової проблеми

Публікації:

- A. Ivaniuk and G. Kriukova, "On Geometric Properties of Adversarial Examples,"2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Cracow, Poland, 2021, pp. 964-967
- Ivaniuk A. Speech audio modeling by means of causal moving average equipped gated attention / A. Ivaniuk // Могилянський математичний журнал. - 2022. - Т. 5. - С. 53-56.

- A. Ivaniuk (2024). ”Latent diffusion model for speech signal processing.”Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems, vol. 61, pp. 43-51.

Наукова (науково-технічна) продукція:

Соціально-економічна спрямованість:

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впровадження не планується

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Крюкова Галина Віталіївна
2. Galyna Kriukova

Кваліфікація: к. ф.-м. н., доц., 01.01.06

Ідентифікатор ORCID ID: 0000-0002-5558-0976

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Києво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, буд. 2, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Родіонов Андрій Миколайович
2. Andriy Rodionov

Кваліфікація: к. т. н., доцент, 05.13.21

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Семенов Володимир Вікторович
2. Vladimir Semenov

Кваліфікація: д. ф.-м. н., професор, 01.05.01

Ідентифікатор ORCID ID: 0000-0002-3280-8245

Додаткова інформація:

Повне найменування юридичної особи: Київський національний університет імені Тараса Шевченка

Код за ЄДРПОУ: 02070944

Місцезнаходження: вул. Володимирська, буд. 60, Київ, 01033, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Олецький Олексій Віталійович
2. Oleksiy Oletsky

Кваліфікація: к. т. н., доц., 05.13.12

Ідентифікатор ORCID ID: 0000-0002-0553-5915

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Києво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, буд. 2, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Швай Надія Олександрівна
2. Nadiya Shvai

Кваліфікація: к. ф.-м. н., доц., 01.01.06

Ідентифікатор ORCID ID: 0000-0001-8194-6196

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Києво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, буд. 2, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Глибовець Микола Миколайович
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Глибовець Микола Миколайович
Відповідальний за підготовку облікових документів	Басенко Олена Едуардівна
Реєстратор	УкрІНТЕІ
Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності	Юрченко Тетяна Анатоліївна