

СТВОРЕННЯ ВЕБ-РЕДАКТОРУ ДЛЯ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ТА ЇХ ЕКСПОРТУ ДЛЯ DEEP LEARNING ФРЕЙМВОРКІВ

Бекузаров М.Р., Спірідонова У.О.

Науковий керівник – ст.викладач кафедри ПІ Широкопетлева М.С.

Харківський національний університет радіоелектроніки
(61166, Харків, преси. Науки, 14, каф. Програмної інженерії,
тел. (057) 702-14-46)

This paper discusses the particular properties of modern architectures of neural networks, that lead to complication in their design and implementation. Modern neural networks have 2 main distinct features – their computation graphs are non-linear and they often consist of repeated blocks. Representing these structures in a programming language is difficult and prone to errors, so a solution to simplify this process was proposed –to develop a graphical editor, which would allow to create, edit, share and even generate code for a given Deep Learning framework. An interactive visual representation of the network's architecture as a 2-D graph should greatly simplify the process of designing a neural network.

Галузь Глибокого Навчання (Deep Learning) є підгалуззю Машинного Навчання і зосереджена на розробці і використанні штучних нейронних мереж. Вона почала формуватися як окрема дисципліна із 2012 року, після виходу в світ відомої статті «Image Net Classification with Deep Convolutional Neural Networks» [1].

У даній роботі розглядаються особливості архітектур сучасних нейронних мереж, що призводять до ускладнення їх проектування, і пропонується рішення – розробка веб-редактору для створення і редагування архітектури нейронних мереж, в основі якого лежить представлення архітектури мережі у вигляді орієнтованого графу.

Нейронна мережа являє собою в простішому випадку послідовність математичних перетворювань (операцій), що очікує на вході вектор дійсних чисел, а на виході – одне число/вектор/матрицю/тензор також дійсних чисел. Ці операції іменуються «шарами» нейронної мережі, і одна операція приймає на вхід результат попередньої, проводить над ним якесь обчислення і передає його далі. При цьому шари нейронної мережі можуть бути організовані не тільки послідовно – мережа може мати два вхідних шара (наприклад, так звані «Сіамські» мережі, що отримують на вхід два зображення і відстежують переміщення якого-небудь об'єкта), два виходи або навіть мати 1 вхід і 1 вихід – але містити нелінійність усередині (так звані «Residual connections» або «Skip connections» [2]). В загальному випадку нейронна мережа являє собою орієнтований граф обчислень, де множиною вершин є множина шарів (операцій) мережі, а множиною ребер – послідовність застосування цих операцій над даними.

За час розвитку Deep Learning було винайдено багато типів шарів (тобто, різних математичних операцій) і багато архітектур (тобто, способів з'єднання цих шарів) нейронних мереж. Основними рисами сучасних архітектур нейронних мереж є наявність нелінійностей у проміжних шарах та велика повторюваність шарів у самій архітектурі (наприклад, майже вся мережа Resnet складається із однакових блоків по 4 шари в кожному, лише з різними параметрами). Архітектури створюються за допомогою сучасних фреймворків (наприклад Keras, pytorch) у вигляді об'єктів у програмному коді. Але програмний код являє собою одновимірну сутність, яка не є придатною до представлення нелінійних сутностей, таких, як нейронна мережа, - що приводить до ускладнення їх опису і збільшення кількості помилок – наприклад, з'єднання шарів мережі в помилковому порядку.

Для запобігання цих проблем було розроблено веб-додаток – редактор для створення архітектур нейронних мереж та експорту їх у готовий програмний код для обраного фреймворка. Цей редактор дозволяє створювати архітектури у вигляді діаграм – додавати / видаляти шари мережі, з'єднувати їх, задавати специфічні для різних типів шарів параметри. Діаграма є набагато простішою для візуального сприйняття ніж програмний код. Параметри задаються за допомогою форми із полями, що відповідають параметрам конкретного шару мережі – це дозволяє зменшити ймовірність помилки (наприклад, завдання параметра, що не підтримується цим шаром). Окремі шари нейронної мережі можна групувати в блоки. Користувач також має можливість згенерувати програмний код для відтворення цієї архітектури в обраному Deep Learning фреймворку мовою Python, обираючи загальні параметри, як-то: символ переносу рядку (Windows чи *nix -style), або специфічні для конкретного фреймворка (наприклад, чи спробувати використовувати Sequential Api в Keras). Користувач також має отримати опис своєї у форматі JSON.

Отже, було розглянуто особливості структури сучасних нейронних мереж та розроблено веб-застосунок з наміром усунути складнощі, що виникають під час їх проектування та створення, а саме – повторюваність частин мережі та представлення нелінійної структури мережі в коді. Цей веб-додаток надає можливість проектування, редагування та експорту архітектур нейронних мереж, з можливістю генерації коду для їх створення для сучасних Deep Learning фреймворків. Представлення архітектури мережі у вигляді діаграми / графу є набагато простішим для сприйняття ніж програмний код, що призводить до значно меншої кількості помилок при розробці.

Перелік посилань

1. <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
2. <https://arxiv.org/abs/1512.03385>