

Training instabilities in neural network-based sequential data modeling

Abstract.

This thesis presents a series of publications regarding instabilities during neural networks' optimization process (also called training) in sequential data modeling. In our study, we investigate two approaches to solving such problems: specialized neural architectures and second-order optimizers.

In the first part of our work, we focus on custom architectures reducing problems with training instabilities. First, we show that a combination of fixed state processing and custom convolutional neural network architecture is sufficient to successfully process a series of medical images. Based on this approach, we introduce the first fully-automatic allergic skin reaction recognition system. We show that our method yields accurate results comparable between patients, as opposed to current, manual state-of-the-art. Next, we propose a minimal gated recurrent neural architecture with an arbitrary, nonlinear state transformation. We theoretically show that this architecture alleviates gradient propagation issues while using fewer parameters than its competitors. Further, we present experimental results which indicate that our method often outperforms state-of-the-art solutions, especially on problems with high state drift. Interestingly, we note that the recurrent neural network's performance degrades with the depth of the state transformation. This observation holds on several tasks for two different architectures.

In the second part, we look closer at second-order optimization algorithms as an alternative to currently often-used gradient ones. To this end, we propose a modification of an evolution strategy algorithm, tailoring it for neural network training. We show that such an approach allows for optimizing both convolutional and recurrent neural networks. Notably, the proposed method can effectively train Jordan recurrent neural networks on tasks with significant gradient propagation problems, in contrast to the state-of-the-art first-order method. Finally, we conclude that combining second-order optimization with gradient information results in the best-performing models.

Overall, our research contributes to the mitigation strategies against training instabilities of neural architectures on sequential data. We introduce two new architectural and one optimization solution to the instability problems. One of the architectural solutions serves as a basis for the novel, real-life medical data analysis system.

Keywords: Sequential Data, Recurrent Neural Networks, Neural Network Training, Optimization

Niestabilności w procesie uczenia sieci neuronowych przy modelowaniu danych sekwencyjnych

Streszczenie.

W niniejszej pracy przedstawiamy serię publikacji poświęconej zagadnieniom niestabilności w procesie uczenia (optymalizacji) sieci neuronowych przy modelowaniu danych sekwencyjnych. W ramach przedstawionych prac badamy dwa podejścia do rozwiązania problemów związanych z niestabilnościami: dedykowane architektury sieci neuronowych oraz optymalizatory drugiego rzędu.

W pierwszej części pracy skupiamy się na architekturach sieci neuronowych zaprojektowanych na potrzeby niwelacji niestabilności w procesie uczenia. Pokazujemy, że połączenie z góry narzuconego, nieelastycznego sposobu przetwarzania stanu oraz dedykowanej konwolucyjnej sieci neuronowej wystarczy aby poprawnie przetwarzać serię obrazów medycznych. Na podstawie tego podejścia wprowadzamy pierwszy w pełni automatyczny system do rozpoznawania reakcji alergicznych skóry. Zaproponowana przez nas metoda otrzymuje dobre wyniki, które mogą być porównywalne pomiędzy pacjentami, w przeciwieństwie do obecnego, manualnego podejścia. Następnie prezentujemy minimalną rekurencyjną sieć neuronową z mechanizmem bramkującym oraz arbitralnym, nieliniowym przekształceniem stanu. Pokazujemy teoretycznie, że zaproponowana architektura rozwiązuje problemy z propagacją gradientu, przy użyciu mniejszej liczby parametrów niż konkurencyjne architektury. Ponadto demonstrujemy wyniki eksperymentalne, które wskazują, że nasz model często uzyskuje lepsze wyniki w porównaniu do obecnie używanych modeli, w szczególności na problemach z dużym dryfem stanu. W wyniku eksperymentów dla dwóch różnych architektur na kilku różnych zadaniach dochodzimy do ciekawego wniosku - jakość wyników rekurencyjnej sieci neuronowej spada wraz z głębokością transformacji stanu.

W drugiej części pracy skupiamy się na optymalizatorach drugiego rzędu jako alternatywie dla obecnie często używanych algorytmów bazujących na gradiencie. W ramach tych prac prezentujemy modyfikację algorytmu strategii ewolucyjnej, którą przystosowujemy do treningu sieci neuronowych. Pokazujemy, że nasz algorytm pozwala na optymalizację konwolucyjnych i rekurencyjnych sieci neuronowych. W szczególności zaproponowana metoda potrafi nauczyć rekurencyjną sieć neuronową Jordana na zadaniach z poważnymi problemami propagacji gradientu, które nie są rozwiązywane przez metody pierwszego rzędu. Pokazujemy również, że łącząc optymalizację drugiego rzędu z informacją o gradiencie otrzymujemy najlepiej działające modele.

Podsumowując, nasza praca przyczynia się do rozszerzenia strategii mitygacji niestabilności treningu sieci neuronowych przy modelowaniu danych sekwencyjnych. Wprowadzamy dwa architektoniczne i jedno optymalizacyjne rozwiązanie problemów ze stabilnością procesu optymalizacji sieci neuronowych. Jedno z architektonicznych rozwiązań stanowi podstawę nowatorskiego systemu analizy danych medycznych.

Słowa kluczowe: Dane Sekwencyjne, Rekurencyjne Sieci Neuronowe, Trening Sieci Neuronowych, Optymalizacja