

Streszczenie

Autor: mgr inż. Grzegorz Rafał Dec, numer albumu: d491

Promotor: dr hab. inż. Zbigniew Hajduk, prof. PRz

Słowa kluczowe: FPGA, ZYNQ, LSTM, BPTT

Celem pracy doktorskiej było przeprowadzanie analizy możliwości implementacyjnych rekurencyjnych sieci neuronowych typu LSTM na układach FPGA z intencją uzyskania porównywalnych lub krótszych czasów obliczeń w stosunku do GPU lub uzyskania wysokiej wydajności obliczeniowej w sytuacjach, gdy GPU lub TPU nie może zostać wykorzystane. W ramach pracy opracowano nową metodę implementacyjną wykorzystującą zarówno układy FPGA, jak i układy hybrydowe z wbudowanym procesorem i matrycą programowalną, zarówno dla zadania klasyfikacji w wybranym procesie przemysłowym, jak i wspierającego proces uczenia sieci LSTM dla wybranego procesu przemysłowego. Prezentowane w ramach pracy podejścia porównano z dotychczas opracowanymi implementacjami.

Zakres pracy obejmuje analizę różnych metod implementacji funkcji aktywacji, ze szczególnym uwzględnieniem dokładności, szybkości działania i wykorzystanych zasobów, zaprezentowanie nowego podejścia implementacyjnego oraz wyselekcjonowanie na drodze eksperymentów różnych wariantów, w tym spotykanych w literaturze, wykorzystywanych w dalszych badaniach. Ponadto obejmuje implementację modułu komórki LSTM z nowym sposobem organizacji obliczeń i prezentację wykorzystania omawianego modułu do celów budowy dowolnej sieci z warstwą LSTM. Dodatkowo, zaprojektowano i zaimplementowano sieć LSTM w strukturach FPGA realizującą zadanie detekcji wadliwego uderzenia w procesie kucia na zimno w oparciu o klasyfikację binarną. Wykonano różne warianty implementacyjne z wykorzystaniem układu FPGA oraz układu z podziałem na sprzęt i oprogramowanie tj. Xilinx ZYNQ. Oprócz tego zakres pracy obejmuje projekt i implementację układu cyfrowego wspierającego proces uczenia sieci LSTM, będących rozszerzeniem wariantów przygotowanych w ramach projektu klasyfikatora do zadania kucia na zimno. Przedstawione w pracy podejścia zostały przeanalizowane pod kątem dostępnych na rynku układów FPGA, uwzględniając analizę kosztów oraz ilość dostępnych zasobów sprzętowych. W ramach pracy wykonano analizę wpływu dokładności aproksymacji funkcji aktywacji na działanie sieci, zarówno w przypadku zadania klasyfikacji, jak i na proces uczenia sieci LSTM.

Uzyskane wyniki pozwalają sądzić, że wykorzystanie układów FPGA oraz układów hybrydowych w celu implementacji sieci LSTM jest wykonalne i może skutkować znaczącym przyspieszeniem obliczeń, nawet w porównaniu z GPU. Możliwe jest również wykonanie układu cyfrowego w strukturach FPGA wspomagającego uczenie sieci LSTM.