

ИЗВЕШТАЈ

о прегледу мастер рада

„Систем за паралелизацију неуронских мрежа на OpenCL акцелераторима“

кандидата Петра Јовановића

Одлуком Наставно-научног већа Математичког факултета која је донета на 322. редовној седници одржаној 9. јуна 2015. године именовани смо за чланове комисије за преглед и оцену мастер рада под насловом „Систем за паралелизацију неуронских мрежа на OpenCL акцелераторима“ кандидата Петра Јовановића, студента мастер студија на студијском програму Информатика на Математичком факултету.

I Област рукописа

У рукопису „Систем за паралелизацију неуронских мрежа на OpenCL акцелераторима“ кандидат Петар Јовановић бави се описом и имплементацијом библиотеке за брзо тренирање великих неуронских мрежа коришћењем графичких картица заснованих на технологији OpenCL. Приликом израде тезе коришћена су знања из области програмирања, вештачке интелигенције и нумеричке математике.

II Структура рукописа и кратак приказ

Рукопис се састоји од 47 страна, организованих у 6 поглавља.

У уводном поглављу кандидат описује област програмирања графичких картица, пре свега платформе CUDA и OpenCL, као и постојеће библиотеке за развој неуронских мрежа, попут библиотека Theano, TensorFlow, Keras и слично.

У поглављу „Аутоматско диференцирање“ уводе се технике аутоматског диференцирања које дају аутоматизацију рачунања градијента и помоћу којих се лакше имплементира алгоритам тренирања неуронских мрежа. Аутоматско диференцирање је врло блиско класичном, симболичком, али оно не садржи конструкцију симболичког израза за извод, већ само директно рачунање његове вредности у датој тачки, што је за примене довољно. Разликују се случајеви аутоматског диференцирања унапред и уназад и у оба се прецизно описују алгоритми аутоматског диференцирања.

У поглављу „Неуронске мреже“ укратко се описују основни појмови неуронских мрежа. Описује се модел перцептрона и вишеслојног перцептрона. Након тога се детаљно описује алгоритам тренирања неуронских мрежа пропагацијом грешака уназад и приказује се његова веза са техникама аутоматског диференцирања. На крају се описују конволутивне неуронске мреже које имају значајну примену код дводимензионалних података, пре свега слика.

Поглавље „Имплементација библиотеке за развој неуронских мрежа на OpenCL-у“ представља централни допринос тезе и у њему се детаљно описује имплементација библиотеке за развој неуронских мрежа у програмском језику Python, а која користи одређен број ефикасних нативних библиотека (на пример, cBLAS за брзо извођење операција над матрицама). Библиотека садржи модул за управљање OpenCL платформом која пружа могућности паралелизације на OpenCL графичким картицама, затим модул за представљање

симболичких израза и њихово аутоматско диференцирање и функције за рад са вишедимензионим низовима на графичкој картици. Приказује се пример употребе библиотеке кроз дефинисање вишеслојног перцептрона и његовог тренирања за препознавање руком писаних цифара (на основу података из колекције MNIST).

У поглављу „Експериментални резултати“ врши се експериментална евалуација библиотеке на примеру тренирања класификационих неуронских мрежа на колекцијама MNIST (која садржи руком писане цифре) и CIFAR (која садржи фотографије објеката из реалног света). Приказују се добијени резултати и уратко се дискутују.

У поглављу „Закључак“ изнети су основни закључци овог рада.

Поглавље „Библиографија“ садржи списак од 42 библиографске јединице које је кандидат користио приликом писања рада.

III Анализа рукописа

У рукопису који смо анализирали кандидат даје јасан преглед основних појмова потребних за имплементацију библиотеке за развој неуронских мрежа, као и опис архитектуре и имплементације саме библиотеке. Предложени систем подржава конструкцију симболичке репрезентације израза са n -димензионалним тензорима, уз аутоматско рачунање градијента функција, што може значајно олакшати имплементацију дубоких неуронских мрежа. Такође, у оквиру рада су имплементиране неке популарне архитектуре неуронских мрежа и њихове перформансе су анализирани над референтним скуповима података (MNIST, CIFAR). Добијени експериментални резултати показују да је помоћу имплементираних библиотека могуће успешно изградити неуронске мреже за класификацију слика и да је време тренинга на графичкој картици мање него на централном процесору. Убрзање је мање него у случају сличних библиотека заснованих на CUDA технологији, тако да постоји простор за даљи рад. Тренутан скуп подржаних оператора није довољан за имплементацију свих постојећих архитектура неуронских мрежа, али постојећи интерфејс омогућава лако проширивање.

IV Закључак и предлог

Описом теоријских неуронских мрежа и имплементацијом библиотеке за њихово тренирање на OpenCL графичким картицама кандидат је приказао задовољавајући степен и теоријског и практичног стручног знања. На основу наведеног Комисија предлаже да се рукопис под насловом:

„Систем за паралелизацију неуронских мрежа на OpenCL акцелераторима“

кандидата Петра Јовановића прихвати као мастер рад и да се закаже његова јавна одбрана.

Београд, 28. 7. 2018.

Комисија:

др Филип Марић, ментор

др Саша Малков

др Младен Николић