

FAKULTETU TEHNIČKIH NAUKA U NOVOM SADU

OBRAZAC ZA PRIJAVU TEHNIČKOG REŠENJA

Obavezni podaci:

Autor/Autori rešenja:

Vuk Vranjković, Rastislav Struharik

Naziv tehničkog rešenja:

Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela

Kategorija tehničkog rešenja:

Softver, M85

Za koga je rešenje rađeno i u okviru kog projekta MNTR:

- Fakultet tehničkih nauka u Novom Sadu
- Projekat tehnološkog razvoja TR-32016

Tehničko rešenje je prihvatio i koristi:

Fakultet tehničkih nauka u Novom Sadu za potrebe daljih istraživanja

Godina kada je rešenje urađeno:

2016.

Kako su rezultati verifikovani i od strane kog tela:

- Razvojem HDL opisa IP jezgara i njihovom verifikacijom kao i proverom performansi na FPGA kolu.
- Rezultate je verifikovalo Naučno-nastavno veće Fakulteta tehničkih nauka u Novom Sadu.

Oblast na koju se tehničko rešenje odnosi:

Hardversko ubrzavanje algoritama u FPGA tehnologiji.

Problemi koji se tehničkim rešenjem otklanjaju ili minimizuju:

Jedinstvena arhitektura realizovana kao IP jezgro u predlogu tehničkog rešenja može da implementira tri najčešće korišćena tipa klasifikatora u oblasti mašinskog učenja: DT (Decision Trees, stabla odluka), ANN (Artificial Neural Networks, veštačke neuronske mreže) i SVM (Support Vector Machine). Do sada predložena hardverska rešenja mogu da implementiraju samo jednu vrstu klasifikatora. Ukoliko je prilikom korišćenja ovih algoritama u nekom sistemu potrebno menjati algoritam predviđanja, tada su postojeća rešenja neprihvatljiva. Dodatno, pošto je predložena arhitektura skalabilna, moguće je jednostavno unapređenje i proširenje postojećeg sistema.

Stanje rešenosti pitanja istog problema u svetu:

Postoji veliki broj objavljenih rezultata u oblasti hardverskih implementacija prediktivnih modela mašinskog učenja. Stabla odluka su mašinski klasifikatori sa jednostavnom idejom koji se primenjuju već duže vreme. Iako se ovi modeli primenjuju veoma često, tek odskora su se pojavile prve hardverske arhitekture za akceleraciju ovih modela. Hardverska implementacija za trening ovih klasifikatora data je u [1]. Implementacija poznatog CART (Classification and Regression Tree) algoritma opisana je u [2]. Paralelna protočna arhitektura za ortogonalna stabla odluka je prikazana u [3]. IP (Intellectual Property) jezgro koje može da realizuje ortogonalna, neortogonalna i nelinearna stabla odluka, pri čemu je pogodno za integraciju u SoC (System on Chip) rešenja opisano je u [4] i [5].

SVM klasifikatori su često korišćeni u industrijskim sistemima. Jedan od problema sa ovim modelima je što su komplikovani za izračunavanje u odnosu na ostale vrste klasifikatora. Zbog navedenih razloga hardverski akceleratori za ove sisteme su se pojavili odmah nakon njihovog objavljivanja u [6]. Hardverska akceleracija trening faze ovih algoritama prikazana je u [7]. Najpoznatiji algoritam za treniranje SVM klasifikatora, SMO (Sequential Minimal Optimization) implementiran je [8] i [9]. Što se tiče samih prediktivnih modela predloženo je nekoliko rešenja. U [10] je predložen generator prediktivnih modela pogodnih za implementaciju na FPGA (Field programmable gate array) čipu. Jedna paralelna arhitektura je prikazana u [11] dok je arhitektura koja koristi radijalni kernel predložena u [12].

ANN su najpoznatiji mašinski klasifikatori kod kojih je hardverska akceleracija u upotrebi više od dve decenije. Postoje analogna, digitalna kao i hibridna hardverska rešenja za akceleraciju ovih klasifikatora. Postoji i nekoliko knjiga pisanih na ovu temu od koji je jedna [13]. U [14] je predložena paralelna arhitektura sa fiksnom prezentacijom brojeva. Multi-Layer Perceptron (MLP) neuronska mreža je realizovana u [15] na FPGA čipu. Mreža je implementirana kao slojevita struktura. Arhitektura neuronske mreže kao SoC rešenja prikazana je u [16].

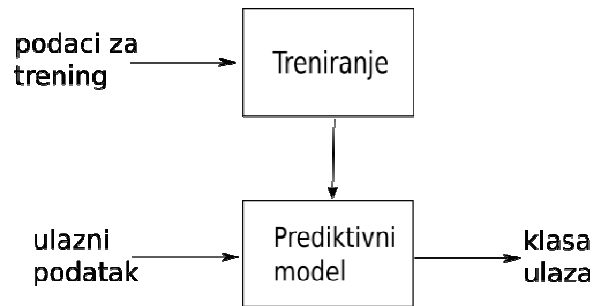
Kao što se može videti postoji velik broj hardverskih rešenja u oblasti mašinskog učenja. Uprkos tome, niti jedna od predloženih arhitektura ne može da implementira više od jednog tipa klasifikaora.

Reference:

- [1] Rastislav JR Struharik and Ladislav A Novak. Evolving decision trees in hardware. *Journal of Circuits, Systems, and Computers*, 18(06): 1033–1060, 2009.
- [2] Grigorios Chrysos, Panagiotis Dagritzikos, Ioannis Papaefstathiou, and Apostolos Dollas. Hc-cart: A parallel system implementation of data mining classification and regression tree (cart) algorithm on a multi- fpga system. *ACM Trans. Archit. Code Optim.*, 9(4):47:1–47:25, January 2013.
- [3] Rastislav JR Struharik and Ladislav A Novak. Hardware implementation of decision tree ensembles. *Journal of Circuits, Systems, and Computers*, 22(05), 2013.
- [4] Rastislav JR Struharik and Ladislav A Novak. Intellectual property core implementation of decision trees. *IET computers & digital techniques*, 3(3):259–269, 2009.
- [5] JR Struharik. Implementing decision trees in hardware. In *Intelligent Systems and Informatics (SISY)*, 2011 IEEE 9th International Symposium on, pages 41–46. IEEE, 2011.
- [6] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [7] Davide Anguita, Andrea Boni, and Sandro Ridella. A digital architecture for support vector machines: theory, algorithm, and fpga implementation. *Neural Networks, IEEE Transactions on*, 14(5):993–1009, 2003.
- [8] Ta-Wen Kuan, Jhing-Fa Wang, Jia-Ching Wang, Po-Chuan Lin, and Gaung-Hui Gu. Vlsi design of an svm learning core on sequential minimal optimization algorithm. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 20(4):673–683, 2012.
- [9] Chih-Hsiang Peng, Po-Chuan Lin, Shovan Barma, Jhing-Fa Wang, Hong-Yuan Peng, Karunanithi Bharanitharan, and Ta-Wen Kuan. Low-power enhanced system-on-chip design for sequential minimal optimisation learning core with tri-layer bus and butterfly-path accelerator. *IET Computers & Digital Techniques*, 2015.
- [10] Davide Anguita, Luca Carlino, Alessandro Ghio, and Sandro Ridella. A fpga core generator for embedded classification systems. *Journal of Circuits, Systems, and Computers*, 20(02):263–282, 2011.
- [11] Christos Kyrkou and Theodoris Theodorides. A parallel hardware architecture for real-time object detection with support vector machines. *Computers, IEEE Transactions on*, 61(6):831–842, 2012.
- [12] V Vranjkovic and Rastislav Struharik. New architecture for svm classifier and its application to telecommunication problems. In *Telecommunications Forum (TELFOR)*, 2011 19th, pages 1543–1545. IEEE, 2011.
- [13] Amos R Omondi and Jagath Chandana Rajapakse. *FPGA implementations of neural networks*, volume 365. Springer, 2006.
- [14] Hirokazu Madokoro and Kazuhito Sato. Hardware implementation of back-propagation neural networks for real-time video image learning and processing. *Journal of Computers*, 8(3):559–566, 2013.
- [15] Antony Savich, Medhat Moussa, and Shawki Areibi. A scalable pipelined architecture for real-time computation of mlp-bp neural networks. *Microprocessors and Microsystems*, 36(2):138–150, 2012.
- [16] Dmitri Vainbrand and Ran Ginosar. Scalable network-on-chip architecture for configurable neural networks. *Microprocessors and Microsystems*, 35(2):152–166, 2011.

Tehnički detalji predloženog rešenja:

Cilj naučne oblasti mašinskog učenja je da se razviju sistemi koji na osnovu ulaznih podataka mogu sami da donesu ispravnu odluku. Ovi sistemi se sastoje od dva dela: dela za formiranje prediktivnog modela i samog prediktivnog modela, kao što je prikazano na slici 1.



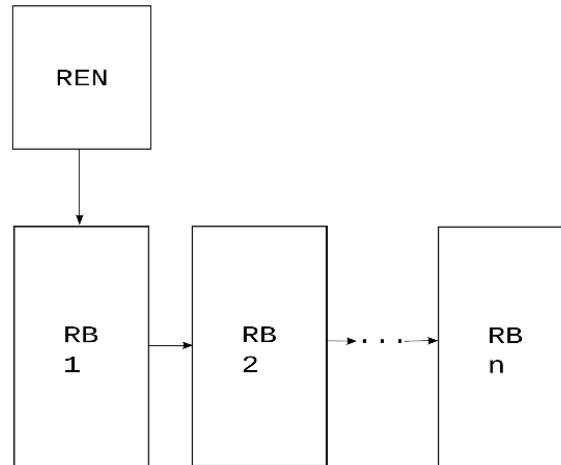
Slika 1: Mašinsko učenje

U zavisnosti od vrste algoritma, prediktivni model se može intepretirati kao znanje ili kao inteligencija. Ulazni podaci se mogu predstaviti kao vektor realnih brojeva, koji će biti označen sa x . Tri najčešće korišćena prediktivna modela, ANN, SVM, DT i njihove varijacije, sadrže veliki broj zajedničkih operacija. Ova činjenica omogućuje da se projektuje jedna univerzalna digitalna rekonfigurabilna arhitektura koja može da radi kao veći broj različitih prediktivnih modela mašinskog učenja. Formule koje arhitektura treba da sračunava su:

$$\begin{aligned} & \mathbf{w} \cdot \mathbf{x} + b \\ & f(\mathbf{w} \cdot \mathbf{x} + b) \\ & \sum_w \alpha f(\mathbf{w} \cdot \mathbf{x}) + b \\ & \sum_w \alpha f(\|\mathbf{w} - \mathbf{x}\|) + b \\ & f(\mathbf{w} \cdot \mathbf{x} + b) \\ & f(\|\mathbf{w} - \mathbf{x}\|) \end{aligned}$$

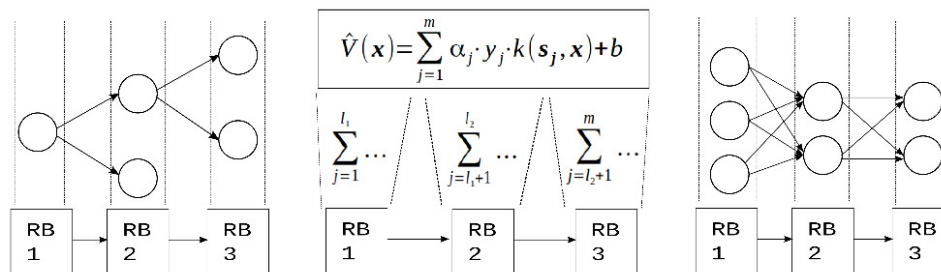
gde $\mathbf{w}$ predstavlja vektor težinskih suma, b je skalarna vrednost, α su Lagranžovi koeficijenti koji se koriste kod SVM klasifikatora, dok je f nelinearna funkcija.

Može se primetiti da sve formule dele slične operacije: skalarni proizvod, nelinearna funkcija, množenje, sabiranje, vektorsko oduzimanje i modulo vektora. Ideja univerzalne arhitekture je da se napravi digitalni sistem, koji može da računa sve navedene operacija, dok se sa posebnim kontrolnim putem kontroliše koje operacije se koriste i kojim redosledom. U nastavku će biti opisana jedna arhitektura koja implementira ovu ideju. Sama arhitektura je upakovana kao IP jezgro koje ćemo zvati RMLC (Reconfigurable Machine Learning Classifier). Arhitektura je organizovana kao lanac istovetnih blokova, kao što je prikazano na slici 2.



Slika 2: Interna struktura RMLC arhitekture

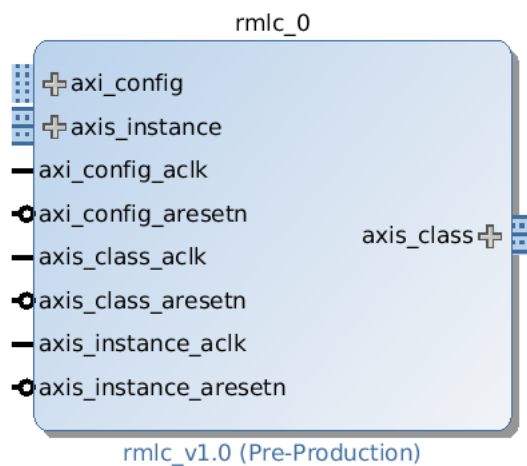
Na slici 2 ti blokovi su označeni sa RB (reconfiguration block). Module REN (reconfiguration enable) služi da selektuje, tokom rekonfiguracije, trenutno aktivni blok. Pre sinteze može se odrediti broj RB jedinica unutar arhitekture. Time je obezbeđena odlična skalabilnost arhitekture, pošto se ona može prilagođavati raspoloživim resursima, pre sinteze.



Slika 3: Način mapiranja različitih klasifikatora na RMLC arhitekturu

Svaki RB blok implementira deo računanja formule svakog od podržanih prediktivnih modela. U zavisnosti kako je IP jezgro konfigurisano, RB blok sračunava različite delove prethodno navedenih formula. Kao što je prikazano na slici 3, može se videti da se RB blok može interpretirati različito u odnosu na trenutnu konfiguraciju arhitekture. U slučaju da se radi o ANN modelima, RB predstavlja jedan sloj neuronske mreže. Ako se radi o SVM klasifikatorima, tada RB blok sračunava deo Lagranžove sume, neophodne za klasifikaciju SVM-ova. Kada je arhitektura konfigurisana da radi kao stablo odluke tada RB blok sračunava neki od čvorova unutar jednog nivoa stabla. Ovakva organizacija arhitekture omogućava dobru skalabilnost i protočnost obrade na nivou arhitekture, a kao posledica toga dolazi i veliko ubrzanje u odnosu na postojeća rešenja. Protočnost obrade je omogućena time što svaki RB blok može da radi u paraleli sa svim ostalim. To znači da dok jedan RB blok obrađuje jednu instancu, prethodni radi na sledećoj itd.

Arhitektura je implementirana na RT nivou abstrakcije i upakovana je u IP jezgro pomoću Xilinx Vivado alata. Izgled samog jezgra unutar IP integratora, koji je jedan od delova Vivado alata, prikazan je na slici 4.



Slika 4: Interfejs RMLC IP jezgra

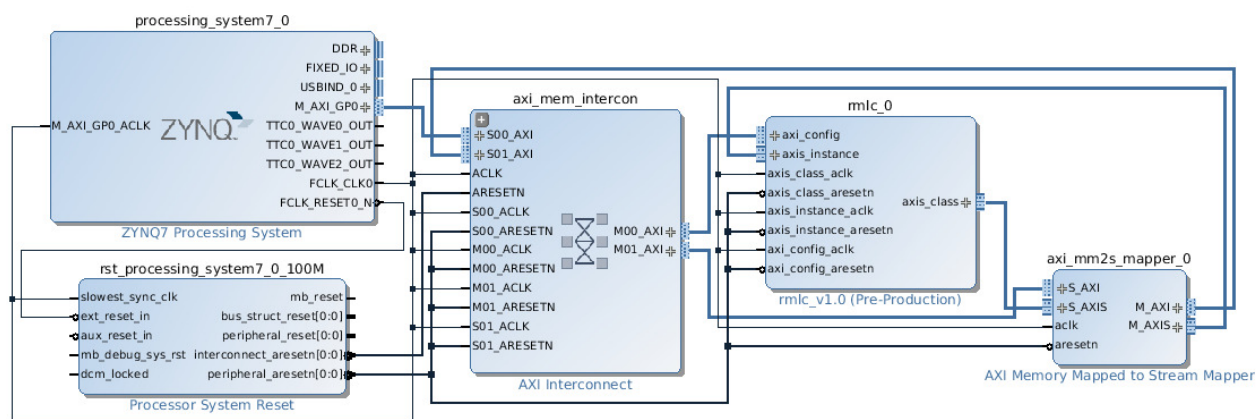
IP jezgro je predviđeno za korišćenje unutar ARM SoC-ova. Stoga su za interfejse odabrane standardne AXI (AMBA eXtensible Interface) magistrale. AXI protokol je deo AMBA (Advanced Microcontroller Bus Architecture) specifikacije i odabran je kao standardni protokol za povezivanje IP jezgara unutar Vivado IP integratora. Unutar Vivado alata podržana su tri standardna AXI protokola: AXI Full, AXI Lite i AXI Stream. Za potrebe RMLC jezgra implementirana su tri interfejsa. Jedan interfejs služi za konfigurisanje RMLC jezgra i implementiran je kao AXI Full protokol. Ovaj interfejs služi za programiranje memorija i registra koji se nalaze unutar RB blokova. Na slici 4 je označen kao *axi_config*. Za ovaj interfejs je odabran AXI Full protokol zato što taj protokol omogućava memorijsko mapiranje i veliku brzinu transfera podataka. Taj protokol je najsloženiji za implementaciju ali je odabran stoga što je za rekonfiguraciju RMLC jezgra potrebno pristupiti pojedinačnim registrima kao i blokovima memorije, a za to je najpogodniji AXI Full protokol.

Druga dva interfejsa implementiraju AXI Stream protokol. Prvi interfejs, pod imenom *axi_instance*, služi za slanje ulazne instance jezgru. Ovaj interfejs je slejv tipa. To znači da ostatak sistema, najčešće procesor, treba da inicira transakciju na njemu. Drugi interfejs koji se zove *axis_class* je master interfejs. RMLC jezgro započinje transakciju na ovom interfejsu. Preko ovog interfejsa se šalje rezultat predviđanja, tj. klasa kojoj je pridružena poslatoj instanci. AXI Stream interfejsi su odabrani za slanje instance i rezultata zato što su najjednostavniji i omogućavaju veliku protočnost podataka.

Xilinx čipovi se mogu podeliti na dve grupe. Jedna grupa sadrži već integrisan procesor. Takvi procesori se zovu hard IP jezgra. U novim Xilinx FPGA familijama ovi procesori su po pravilu ARM-ova hard IP jezgra. Trenutno aktuelni čipovi sa hard IP procesorom pripadaju Zynq familiji. Druga grupa ne sadrži procesor, već se on može realizovati na programabilnom delu čipa. Ovakvi procesori se nazivaju i soft IP jezgrima. U okviru Xilinx-ovih već postojećih IP jezgara postoji Microblaze procesor koji je veoma rekonfigurabilan. Ovaj procesor je najčešće korišćen u FPGA sistemima bez hard IP procesora. Trenutno aktuelne familije

bez hard IP procesora su Artix, Kintex i Virtex. U narednim paragrafima će biti opisano kako se RMLC jezgro može ugraditi u SoC u zavisnosti da li se koristi Zynq AMR procesor ili Microblaze.

Na slici 5 prikazan je jedan od načina povezivanja RMLC jezgra u sistem sa Xilinx Zynq podsistemom.



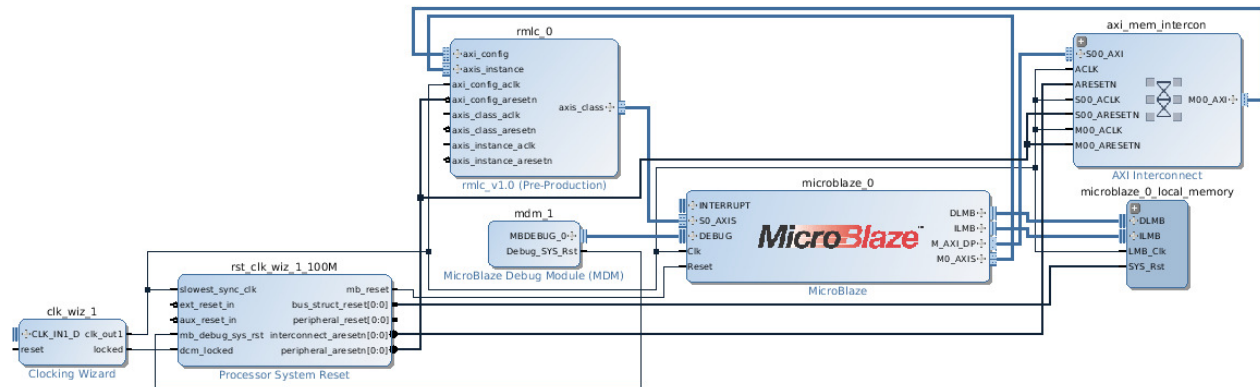
Slika 5: Povezivanje Zynq PS i RMLC IP jezgra

Prikazani sistem služi da ilustruje jedan od mogućih načina kako RMLC jezgro može da se koristi u SoC-u i svakako nije jedini način. AXI Interconnect jezgro je postavljeno između Zynq PS (Processing Systema) i RMLC jezgra. Ovo jezgro se nalazi kao standardna komponenta u Xilinx katalogu IP jezgara i služi da za sve AXI periferije koje se nalaze u Zynq SoC-u obezbedi jedinstven adresni prostor, pomoću kog PS može da im pristupi. Drugim rečima, ono omogućava memorijsko mapiranje komponenti. Na *S00_AXI* slave interfejs AXI Interconnect-a je vezana PS AXI magistrala, preko koje procesor šalje sve potrebne podatke raznim delovima SoC-a. Na *M00_AXI* master interfejs je povezano RMLC jezgro. Kada PS pošalje neki podatak AXI Interconnect-u koji se poklapa sa adresom na koju je mapirano RMLC jezgro, ovaj interfejs će inicirati transakciju i odgovarajući podatak će se poslati RMLC jezgru ili će se pročitati iz njegovih registara.

Druga periferija koja je povezana na AXI Interconnect je AXI Memory Mapped to Stream Mapper jezgro. To je još jedna od standardnih komponenti koje se nalaze u Xilinx-om katalogu. Interfejs *M01_AXI* je upotrebljen da se jezgro poveže u SoC. Ovo jezgro konvertuje AXI lite/full protokol u AXI stream protokol. Ova komponenta se često koristi za povezivanje IP jezgara sa AXI Stream interfejsom u Zynq SoC. Kada se upotrebi ova komponenta, AXI Stream interfejsu, koji nema koncept adrese, može se pristupiti preko PS-a. AXI Memory Mapped to Stream Mapper će primiti jednu ili više ulaznih instanci pa će ih proslediti RMLC jezgru. Kada dobije rezultate od RMLC jezgra proslediće ih natrag PS-u.

Kada se pogleda način na koji je povezano RMLC jezgro u Zynq SoC, može delovati da je bolji izbor za slanje instanci i rezultata bio AXI Lite protokol, pošto u tom slučaju ne bi bio potreban dodatni konvertor. Zynq PS radi na višestruko većoj frekvenciji od ostatka programabilne logike, PL (Programable Logic) FPGA čipa. Stoga slanje instance i primanje rezultata nije kritično u ovakvom sistemu. U slučaju da je RMLC jezgro potrebno ugraditi u SoC na Xilinx FPGA čipu bez integrisanog procesora, kao što su na

primer Kintex i Virtex familija čipova, tada se pokazuje opravdanost izbora AXI Stream protokola za ova dva interfejsa. Primer povezivanja RMLC jezgra sa Xilinx Microblaze soft procesorom je prikazano na slici 6.



Slika 7: Povezivanje Microblaze procesora i RMLC jezgra

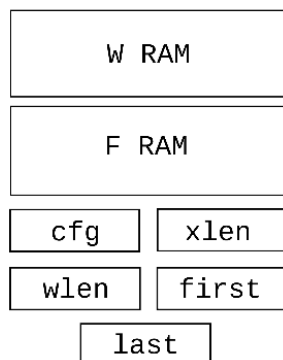
Microblaze procesor se realizuje u PL delu FPGA čipa, kao i ostale periferije. Može da radi na frekvencijama koje su približne frekvencijama rada ostalih periferija. Stoga je u ovom slučaju brzina transfera bitna. Ono što je razlika u odnosu na Zynq SoC je veća mogućnost rekonfigurisanja Microblaze procesora. Na ovaj procesor može se dodati direktno interfejs koji poštuje AXI stream protokol, pa je IP jezgra sa ovim protokolom moguće povezati direktno na Microblaze procesor. Broj AXI stream interfejsa na Microblaze procesoru moguće je podešavati. Za ovu ilustraciju je uzet samo jedan master interfejs i jedan slejv interfejs, što omogućava lako povezivanje RMLC jezgra u SoC. *M0_AXIS* interfejs je povezan na AXI Stream slejv interfejs *axis_instance*, dok je *S0_AXIS* slejv interfejs procesora povezan na *axis_class* master interfejs jezgra. Kao i u slučaju ilustracije Zynq SoC-a, *axi_config* interfejs RMLC jezgra je preko AXI Interconnect-a povezano na AXI master magistralu procesora, koja se u ovom slučaju zove *M_AXI_DP*.

Određivanje adrese za RMLC jezgro se podešava pomoću alata Address Editor unutar Vivado IP integratora. Na slici 7 je prikazan jedan primer dodeljene adrese za RMLC jezgro u slučaju Zynq SoC primera. Adresa koja je dodeljena *axi_config* interfejsu je 0x7AA0_0000. U softveru koji treba da se razvije sa ovaj SoC, kada god se pristupa ovoj adresi ili nekoj adresi koja je do 64k veća od nje, informacije će teći ka ovom interfejsu. Ukoliko se pristupi adresi 0x7AA0_0001, informacije će biti upisane u registar koji je mapiran na adresi 1 unutar samog IP-jezgra.

Cell	Slave Interface	Base Name	Offset Address	Range	High Address
processing_system7_0					
Data (32 address bits : 0x40000000 [1G])					
rmlc_0	axi_config	axi_config_mem	0x7AA0_0000	64K	0x7AA0_FFFF
axi_dma_0	S_AXI_LITE	Reg	0x4040_0000	64K	0x4040_FFFF
axi_dma_0					

Slika 6: RMLC IP dodela adrese

Programski model RMLC arhitekture organizovan je po RB modulima. Svaki RB modul ima baznu adresu na koju se dodaje vrednost adrese odgovarajućeg registra ili memorijske lokacije. Prvi RB modul u nizu ima baznu adresu 0, drugi ima baznu adresu 1 itd. Svaki RB modul ima svoje registre kao i memorijske elemente. Programski model svakog od RB blokova unutar RMLC IP jezgra prikazan je na slici 8.



Slika 8: Programski model RB modula

Bazna adresa svakog od RB modula množi se sa brojem koji je stepen broja 2. Tačna vrednost ovog množitelja je jedan od parametara arhitekture i podešava se pre sinteze dizajna. Na baznu adresu se dodaje vrednost adrese registra iz koga je potrebno čitati ili upisati vrednost. Adrese samih registara su postavljene kao konstante i mogu se menjati pre same sinteze. Na primer, ukoliko je vrednost množitelja bazne adrese 1024 i želimo da čitamo vrednost registra *cfg*, čija je adresa postavljena na vrednost 2, iz RB modula na poziciji 3, tada je efektivna adresa ovog registra 2050. Treći RB modul u nizu ima baznu adresu dva, koju se množi sa 1024 i na to se dodaje vrednost adrese *cfg* registra.

Kao što se može videti, RB moduli imaju 5 registara i 2 memorije. Registri RB modula se zovu: *cfg*, *xlen*, *wlen*, *first* i *last*. Registar *cfg* određuje u kom režimu radi trenutno RB modul. Na primer, ukoliko se u ovom registru nalazi vrednost 2, tada RB modul radi na sračunavanju sume u SVM klasifikatoru sa polinomijalnim kernelom. Registar *xlen* određuje dužinu ulazne instance. Registar *wlen* ima različite funkcije u zavisnosti od toga kako je RB modul konfigurisan. Ovaj registar i ne mora da se koristi u nekim vrstama klasifikatora. Ako RB modul radi kao SVM klasifikator, tada ovaj registar određuje broj vektora podrške za koje se suma računa unutar modula. Kada RB modul radi kao neuronska mreža tada ovaj registar određuje broj neurona koje RB modul simulira. Registri *first* i *last* su dužine 1 bit i određuju da li je RB modul, prvi ili poslednji u lancu RB modula.

Uz 5 registara, svaki od RB modula ima i dve memorije: W RAM i F RAM. Memorija W RAM sadrži vrednosti težinskih vektora i dodatne potrebne informacije za predikciju. Koji su to tačno vektori i koliko ih ima zavisi od vrste klasifikatora. Memorija F RAM sadrži odbirke nelinearnih funkcija potrebnih za proračun različitih klasifikatora.

Sadržaj W RAM memorije u slučaju kada RB modul radi kao stablo prikazan je na slici 9.

W11	W12	W13	...	W1n	b1	tr1	adr ₁₁	adr ₁₂
W21	W22	W23	...	W2n	b2	tr2	adr ₂₁	adr ₂₂
W31	W32	W33	...	W3n	b3	tr3	cls ₁	cls ₂
...			...					
Wm1	Wm2	Wm3	...	Wmn	bm	trm	adr _{m1}	adr _{m2}

Slika 9: Sadržaj W RAM memorije za stable odluke

Vrednosti W_{xy} predstavljaju težinske vektore neophodne za proračun u različitim čvorovima slojeva stabla. Vrednost b_x je dodatna vrednost neophodna kod nekih vrsta stabala odluka. Vrednost tr_x predstavlja prag. Ukoliko vrednost proračuna bude iznad ove granice u narednom RB bloku proračun se počinje od adrese adr_{x1} , dok u suprotom počinje od adrese adr_{x2} . Ukoliko se na kraju proračuna dobije vrednost koja je kodovana kao klasa, cls_x , tada je proračun završen i rezultat klasifikovanja je cls_x .

Kada RB modul radi kao SVM prediktivni model tada je sadržaj u memoriji prikazan na slici 10. Vrednosti W_{xy} predstavljaju vektore podrške dok su vrednosti α_x Lagranžovi množitelji.

S11	S12	S13	...	S1n	α_1
S21	S22	S23	...	S2n	α_2
S31	S32	S33	...	S3n	α_3
...			...		
Sm1	Sm2	Sm3	...	Smn	α_m

Slika 10: Sadržaj W RAM memorije za SVM

U slučaju neuronskih mreža, sadržaj W RAM memorije je prikazan na slici 11. Vrednosti W_{xy} predstavljaju težinske vektore ili centralni vektor u neuronima. Ukoliko neuronske mreže koriste prag, onda se on čuva u vrednostima b_x .

W11	W12	W13	...	W1n	b1
W21	W22	W23	...	W2n	b2
W31	W32	W33	...	W3n	b3
...			...		
Wm1	Wm2	Wm3	...	Wmn	bm

Slika 11: Sadržaj W RAM memorije za neuronske mreže

Efikasnost RMLC jezgra je poređena sa R Project softverskim paketom. R Project je jedan od dva najkorišćenija softvera, otvorenog koda u oblasti mašinskog učenja. Softver se izvršavao na PC računaru u kom se nalazi AMD Phenom II 1090T procesor koji radi na učestanosti 3.2Ghz. RMLC jezgro je podešeno da radi na učestanosti 113 Mhz. Izvršeno je ukupno 18 standardnih testova iz UCI baze testova. Korišćeni

su sledeći testovi: Breast Cancer (bc), Wisconsin Breast Cancer (bcw), Credit Approval (ca), German Credit (cg), Horse Colic (col), Congressional Voting Records (cvr), Heart Disease (hd), Hepatitis (hep), Haberman's Survival (hs), Ionosphere (ion), Liver Disorders (ld), Labor Relations (lr), Mushrooms (mus), Pima Indians Diabetes (pid), King Rock vs King Pawn (rvp), Spambase (sb), Sonar, Mines vs Rocks (son), Tic-Tac-Toe (ttt).

Svih 18 testova je sprovedeno nad 5 vrsta mašinskih klasifikatora. Korišćena su ortogonalna stabla (AP-DT), SVM modeli sa polinomijalnim kernelom (SVM-P), SVM modeli sa radijalnim kernelom (SVM-R), neuronske mreže MLP (Multilayer Perceptron) tipa (MLP)-ANN i neuronske mreže koje koriste radijalne funkcije (RB-ANN).

Tabela 1: Rezultati testova

Skup	AP-DT	SVM-P	SVM-R	MLP-ANN	RB-ANN
bc	17.489	4.597	6.526	5.125	2.597
bcw	11.448	4.613	9.917	15.868	7.166
ca	9.672	5.454	9.45	3.86	1.696
cg	9.514	7.655	9.035	3.899	1.965
col	11.776	7.54	11.385	5.847	2.924
cvr	19.687	4.882	8.69	8.011	3.711
hd	19.694	5.726	11.263	10.641	4.858
hep	10.917	11.397	17.597	11.568	5.244
hs	6.781	4.006	7.248	7.901	3.633
ion	5.009	4.14	6.218	5.607	2.637
ld	5.651	3.819	8.048	1.347	0.664
lr	10.571	5.861	8.959	6.664	3.268
mus	17.146	4.299	4.479	29.481	13.676
pid	7.112	3.663	6.998	8.986	4.069
rvp	21.293	9.559	12.225	23.274	10.356
sb	4.26	3.747	4.117	4.486	2.317
son	4.774	3.486	4.075	6.086	2.906
ttt	4.204	3.748	5.857	3.947	1.829
Prosek	10.94	5.46	8.45	9.03	4.2

U tabeli 1 prikazana su ubrzanja RMLC arhitekture u odnosu na čisto softversku implementaciju, za odabranih 18 UCI testova. Kao što se može videti, RMLC jezgro daje značajno ubrzanje u odnosu na R Project softverski paket. Pri tome samo jezgro radi na značajno manjoj učestanosti u odnosu na PC procesor. Ovo pokazuje efikasnost RMLC arhitekture. I pored skoro 30 puta manje učestanosti ostvaruju se i do 20 puta veća ubrzanja. U ovim eksperimentima je poređena RMLC arhitektura sa PC softverom. U embedded aplikacijama, gde se koriste procesori sa značajno lošijim performansama, RMLC arhitektura bi mogla da pruži još veća ubrzanja, uz značajno smanjenje potrošnje.

Realizacija rešenja i mogućnost primene:

Rešenje je realizovano kao IP jezgro u jeziku za opis digitalnog hardvera (Verilog), korišćenjem RTL metodologije. Stoga, ovo jezgro se može upotrebiti kao hardverski akcelerator u SoC rešenjima raznih tehnologija (ASIC, FPGA), bez dodatnih modifikacija.

Podnosilac prijave



УНИВЕРЗИТЕТ
У НОВОМ САДУ



ФАКУЛТЕТ
ТЕХНИЧКИХ НАУКА

Трг Доситеја Обрадовића 6, 21000 Нови Сад, Република Србија
Деканат: 021 6350-413; 021 450-810; Централа: 021 485 2000
Рачуноводство: 021 458-220; Студентска служба: 021 6350-763
Телефакс: 021 458-133; e-mail: ftndean@uns.ac.rs

ИНТЕГРИСАНИ
СИСТЕМ
МЕНАџМЕНТА
СЕРТИФИКОВАН ОД:



Наш број: 01-сл

Ваш број:

Датум: 2016-11-03

ИЗВОД ИЗ ЗАПИСНИКА

Наставно-научно веће Факултета техничких наука у Новом Саду, на 25. редовној седници одржаној дана 26.10.2016. године, донело је следећу одлуку:

-непотребно изостављено-

ТАЧКА 10. Питања научноистраживачког рада и међународне сарадње

Тачка 10.2.3: У циљу верификације новог техничког решења усвајају се рецензенти:

- Проф. др Теуфик Токић, Електронски факултет у Нишу
- Доц. др Татјана Николић, Електронски факултет у Нишу

Назив техничког решења:

**"РЕКОНФИГУРАБИЛНО IP ЈЕЗГРО ЗА АКЦЕЛЕРАЦИЈУ DT, SVM и ANN
ПРЕДИКТИВНИХ МОДЕЛА"**

Аутори техничког решења: Вук Врањковић, Растислав Струхарик.

-непотребно изостављено-

Записник водила:

Јасмина Димић, дипл. правник

Тачност података оверава
Секретар

Иван Нешковић, дипл. правник

Декан



Проф. др Раде Дорословачки

Softver:

Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela

Rukovodilac projekta: prof. dr Ljiljana Živanov

Odgovorno lice: dr Vuk Vranjković

Autori: Vuk Vranjković, Rastislav Struharik
Fakultet tehničkih nauka (FTN), Novi Sad

Razvijeno: u okviru projekta tehnološkog razvoja TR-32016

Godina: 2016.

Primena: novembar 2016.

Kratak opis

Ukoliko je prilikom projektovanja sistema neophodno veoma kratko vreme za izračunavanje rezultata prediktivnih modela mašinskog učenja tada hardverski akceleratori mogu biti jedino prihvatljivo rešenje za implementaciju sistema. Dodatno, ukoliko postoje ograničenja u potrošnji koju sistem ima na raspolaganju, hardverski akcelerator može biti od pomoći. Ova ograničenja postoje kod savremenih embedded sistema i SoC (System on Chip) rešenja. Razvijeno IP jezgro za akceleraciju prediktivnih modela mašinskog učenja olakšava izgradnju embedded sistema i SoC rešenja u kojima je neophodno mašinsko klasifikovanje.

Tehničke karakteristike:

Rešenje je realizovano kao IP jezgro u jeziku za opis digitalnog hardvera (Verilog) korišćenjem RTL metodologije. Stoga, ovo jezgro se može upotrebiti kao hardverski akcelerator u SoC rešenjima raznih tehnologija (ASIC, FPGA), bez dodatnih modifikacija.

Tehničke mogućnosti:

Jedinstvena arhitektura realizovana kao IP jezgro u predlogu tehničkog rešenja može da implementira tri najčešće korišćena tipa klasifikatora u oblasti mašinskog učenja: DT (Decision Trees, stabla odluka), ANN (Artificial Neural Networks, veštačke neuronske mreže) i SVM (Support Vector Machine). Predložena arhitektura je skalabilna i moguće je jednostavno menjanje trenutnog prediktivnog modela mašinskog učenja.

Realizator:

Fakultet tehničkih nauka – FTN

Korisnik:

Fakultet tehničkih nauka – FTN, Novi Sad

Podtip rešenja:

Softver – M85

Mišljenje

Tehničko rešenje "Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela" autora doc. dr Vuk Vranjkovića, i van. prof. dr Rastislava Struharika, realizovano 2016. godine, prikazano na 12 stranica A4 formata, grupisano je u ukupno četiri glavne celine:

1. Problemi koji se tehničkim rešenjem otklanjaju ili minimizuju,
2. Stanje rešenosti pitanja istog problema u svetu,
3. Tehnički detalji predloženog rešenja,
4. Realizacija rešenja i mogućnost primene.

Tehničko rešenje pripada polju tehničko-tehnoloških nauka i oblasti elektrotehničkog inženjerstva. Naručilac tehničkog rešenja je Fakultet tehničkih nauka u Novom Sadu, Republika Srbija, koji je i korisnik tehničkog rešenja.

Tehničko rešenje je realizovano u okviru projekta "Nove generacije ugrađenih elektronskih komponenti i sistema u neorganskim i organskim tehnologijama za uređaje široke potrošnje" (Broj projekta TR 32016, Program istraživanja u oblasti tehnološkog razvoja za period 2011-2014., Tehnološka oblast - Elektronika, telekomunikacije i informacione tehnologije, Rukovodilac projekta: dr Ljiljana Živanov, redovni profesor).

Na osnovu analize tehničkog rešenja "Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela" autora doc. dr Vuk Vranjkovića, i van. prof. dr Rastislava Struharika, mogu se izvesti sledeći zaključci:

1. Dokumentacija tehničkog rešenja jasno prikazuje kompletnu strukturu tehničkog rešenja – opis problema, daje detaljniji osvrt na stanje u svetu, sadrži odgovarajući prikaz teorijskih osnova na kojima je zasnovano tehničko rešenje i posebno detaljno prikazuje strukturu i primenu realizovanog tehničkog rešenja.
2. Predloženo tehničko rešenje, "Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela", predstavlja efikasan alat za rešavanje problema u oblasti hardverske akceleracije prediktivnih modela mašinskog učenja.
3. Tehničko rešenje predstavlja originalan naučni doprinos sa praktičnom dimenzijom. Predložena arhitektura je rekonfigurabilna, pa je njena primena fleksibilna i univerzalna.

Na osnovu prethodnog, predlažem da se "Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela", autora doc. dr Vuk Vranjkovića, i van. prof. dr Rastislava Struharika, prihvati kao novo tehničko rešenje i u skladu sa Pravilnikom o postupku i načinu vrednovanja, i kvantitativnom iskazivanju naučnoistraživačkih rezultata istraživača ("Službeni glasnik RS", broj 38/2008) klasifikuje kao rezultat "M85 Prototip, nova metoda, softver, standardizovan ili atestiran instrument, nova genska proba, mikroorganizmi".



Prof. Dr Teufik Tokić,
Univerzitet u Nišu
Elektronski fakultet

Softver:

Rekonfigurabilno IP jezgro za akceleraciju DT, SVM i ANN prediktivnih modela

Rukovodilac projekta: prof. dr Ljiljana Živanov

Odgovorno lice: dr Vuk Vranjković

Autori: Vuk Vranjković, Rastislav Struharik

Fakultet tehničkih nauka (FTN), Novi Sad

Razvijeno: u okviru projekta tehnološkog razvoja TR-32016

Godina: 2016.

Primena: novembar 2016.

Kratak opis

Ukoliko je prilikom projektovanja sistema neophodno veoma kratko vreme za izračunavanje rezultata prediktivnih modela mašinskog učenja tada hardverski akceleratori mogu biti jedino prihvatljivo rešenje za implementaciju sistema. Dodatno, ukoliko postoje ograničenja u potrošnji koju sistem ima na raspolaganju, hardverski akcelerator može biti od pomoći. Ova ograničenja postoje kod savremenih embedded sistema i SoC (System on Chip) rešenja. Razvijeno IP jezgro za akceleraciju prediktivnih modela mašinskog učenja olakšava izgradnju embedded sistema i SoC rešenja u kojima je neophodno mašinsko klasifikovanje.

Tehničke karakteristike:

Rešenje je realizovano kao IP jezgro u jeziku za opis digitalnog hardvera (Verilog), korišćenjem RTL metodologije. Stoga, ovo jezgro se može upotrebiti kao hardverski akcelerator u SoC rešenjima raznih tehnologija (ASIC, FPGA), bez dodatnih modifikacija.

Tehničke mogućnosti:

Jedinstvena arhitektura realizovana kao IP jezgro u predlogu tehničkog rešenja može da implementira tri najčešće korišćena tipa klasifikatora u oblasti mašinskog učenja: DT (Decision Trees, stabla odluka), ANN (Artificial Neural Networks, veštačke neuronske mreže) i SVM (Support Vector Machine). Predložena arhitektura je skalabilna i moguće je jednostavno menjanje trenutnog prediktivnog modela mašinskog učenja.

Realizator:

Fakultet tehničkih nauka – FTN

Korisnik:

Fakultet tehničkih nauka – FTN, Novi Sad

Podtip rešenja:

Softver – M85

Mišljenje

Fakultet tehničkih nauka je razvio IP jezgro koje može da implementira DT, SVM i ANN mašinke prediktivne modele. IP jezgra su opisana korišćenjem Verilog jezika za modelovanje hardvera. Arhitektura je modelovana na tehnološki nezavistan način. Pri tome, opisana arhitektura je skalabilna. Zbog dva navedena razloga, IP jezgro se može lako prilagoditi potrebama trenutne aplikacije, pa se stoga može koristiti u širokom spektru embedded i SoC rešenja.

U predloženom tehničkom rešenju razmatran je problem hardverske implementacije prediktivnih modela mašinskog učenja. Analizom postojećih rešenja utvrđeno je da do sada objavljena rešenja ne mogu da implementiraju sva tri najčešće korišćena klasifikatora: DT, SVM i ANN.

Predložena arhitektura je nazvana RMLC (Reconfigurable Machine Learning Classifier). Arhitektura se sastoji od niza istovetnih rekonfigurabilnih blokova, RB (Reconfiguration Block). RB blokovi mogu da rade nezavisno, svaki na zasebnoj ulaznoj instanci. Arhitektura organizovana na prikazani način omogućava veliku protočnost podataka. Kao posledica toga dobija se veliko ubrzanje procesa klasifikacije.

Opisano je na koji način arhitektura može da se poveže u SoC sisteme. Kao primer korišćene su Xilinx FPGA komponente. Razmatrane su komponente koje poseduju hard IP ARM procesor i one koje poseduju samo rekonfigurabilnu logiku. U oba slučaja prikazan je po jedan slučaj SoC rešenja u kojima se koristi RMLC IP jezgro. Objašnjeno je kako se jezgro može povezati sa hard IP ARM procesorom kao i soft IP Microblaze procesorom.

Potom je opisan programski model RMLC IP jezgra. Objašnjeno je koji registri i memorije se nalaze unutar RB blokova. Opisano je čemu registri služe i opisana je struktura memorija u slučajevima kada se arhitektura koristi kao DT, SVM i ANN prediktivni model.

Nakon opisa arhitekture, opisa načina povezivanja arhitekture u Xilinx FPGA čipove i opisa programskog modela arhitekture, prikazani su i eksperimentalni rezultati. Eksperimenti su sprovedeni nad 18 skupova iz UCI baze i prikazana su ubrzanja koja arhitektura pruža u odnosu na R-project softverski alat koji se izvršava na standardnom PC računaru. Eksperimenti su pokazali da predloženo IP jezgro pruža značajna ubrzanja u odnosu na softverski paket R-project.

U skladu sa gore iznetim činjenicama tehničko rešenje ispunjava uslove da bude priznato kao softver (odnosno M85 u skladu sa Pravilnikom o postupku i načinu vredovanja i kvantitativnom iskazivanju naučnoistraživačkih rezultata istraživača, Sl. gl. RS br. 38/08).

Dr Tatjana Nikolić

Vanredni profesor Elektronskog fakulteta, Niš

T. Nikolić



УНИВЕРЗИТЕТ
У НОВОМ САДУ



ФАКУЛТЕТ
ТЕХНИЧКИХ НАУКА

Трг Доситеја Обрадовића 6, 21000 Нови Сад, Република Србија
Деканат: 021 6350-413; 021 450-810; Централна: 021 485 2000
Рачуноводство: 021 458-220; Студентска служба: 021 6350-763
Телефакс: 021 458-133; e-mail: ftndean@uns.ac.rs

ИНТЕГРИСАНИ
СИСТЕМ
МЕНАџМЕНТА
СЕРТИФИКОВАН ОД:



Наш број: 01.сл

Ваш број:

Датум: 2016-12-07

ИЗВОД ИЗ ЗАПИСНИКА

Наставно-научно веће Факултета техничких наука у Новом Саду, на 26. редовној седници одржаној дана 30.11.2016. године, донело је следећу одлуку:

-непотребно изостављено-

ТАЧКА 11. Питања научноистраживачког рада и међународне сарадње

Тачка 11.3.: На основу позитивног извештаја рецензената верификује се техничко решење (M85) под називом:

"РЕКОНФИГУРАБИЛНО IP ЈЕЗГРО ЗА АКЦЕЛЕРАЦИЈУ DT, SVM и ANN ПРЕДИКТИВНИХ МОДЕЛА"

Аутори техничког решења: Вук Врањковић, Растислав Струхарик.

-непотребно изостављено-

Записник водила:

Јасмина Димић, дипл. правник

Тачност података оверава:
Секретар

Иван Нешковић, дипл. правник

Декан



Проф. др Раде Дорословачки