

Ե.13.04 «Հաշվողական մեքենաների, համալիրների, համակարգերի և ցանցերի մաթեմատիկական և ծրագրային ապահովում» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ներկայացված Դավիթ Բունիայանի «Կենսաբժշկական պատկերների խորը ուսուցման տարաբաշխված ամպային մեթոդների մշակում» թեմայով ատենախոսության վերաբերյալ

Դ. Բունիայանի ատենախոսությունը նվիրված է կենսաբժշկական մեծ պատկերների մշակման միջավայրերի համար արդյունավետ խորը ուսուցման տարաբաշխված ամպային (ԽՈՒՏԱ) մեթոդների մշակմանը և գնահատմանը: Այդ նպատակին հասնելու համար հեղինակի ձևակերպմամբ դիտարկված են հետևյալ խնդիրները՝ հետազոտել մեքենայական ուսուցման համակարգերի լայնածավալ տվյալների մշակումը կենսաբժշկական մեծ պատկերների վերադրման խնդրի շուրջը: Աշխատանքում ուսումնասիրվել են մեծածավալ մեքենայական ուսուցման կիրառման խնդիրները, այդ թվում՝ մեծաքանակ տվյալների պահպանումն ու տեղաբաշխումը, գույզահեռացված մոդելների ուսուցումն ու խոր մոդելների հաշվարկումը CPU-ների և GPU-ների կիրառմամբ, ինչպես նաև նրանց արդյունավետության համեմատական վերլուծությունը: Աշխատանքում դիտարկվել է նաև թույլ վերահսկելի եղանակով մոդելների օպտիմիզացիան կենսաբժշկական մեծ նկարների վերադրման խնդրի համար:

Աշխատանքի հետ մանրամասն ծանոթանալուց հետո կարծում եմ, որ անհրաժեշտ է ճշգրտել դիտարկված նպատակը և խնդիրները:

Մեր պատկերացումով աշխատանքի բուն նպատակն է կառուցել ԽՈՒՏԱ մեթոդների գներացման և արդյունավետության վերլուծության միջավայր/ հարթակ, որը բացի նրանից, որ ապահովում է մեթոդների արդյունավետության միասնական վերլուծությունը ու համեմատումը, նաև հնարավորություն է տալիս, կատարելով պարամետրերի արժեքների որոշակի փոփոխություններ, համակցել/ կոմբինացնել արդեն գոյություն ունեցող մեթոդները նոր մեթոդի շրջանակում: Սա իր հերթին բերում է, փաստորեն, աղբյուր մեթոդների առավելությունների միավորմանը մեկ մեթոդի մեջ: Այսպիսով, ապահովվում է մեթոդների կատարելագործման եղանակ ամեն նոր խնդիրը դիտարկելու ժամանակ:

Որպես մոդելային խնդիր միջավայրի համար, աշխատանքում դիտարկված է կենսաբժշկական մեծ պատկերների վերադրման խնդիրը, որին կարելի է հետագայում ավելացնել նոր խնդիրներ՝ առանց միջավայրի որևէ էական փոփոխության:

Հայտնի է, որ մեծ քանակի տվյալների հաշվարկումը պարունակում է տարբեր մարտահրավերներ նույնիսկ օգտագործելով ժամանակակից գերհամակարգիչներ և ամպային տեխնոլոգիաները: Հեղինակի դիտարկված խնդրում կենսաբժշկական մեծ պատկերները ունեն երեք առանձնահատկություններ՝ նրանք իզոտոպային են, ծավալային են (100,000x100,000x20,000 պիքսել) և կարող են հասնել պետաբայթ ծավալի (10⁶ գիգաբայթ):

Իզոտոպային պատկերի դեպքում տարբեր տեղամասերում պատկերի մշակման ալգորիթմը մնում է նույնը: Այդ հնարավորություն է տալիս մեծ պատկերը բաժանել կտորների և գույզահեռ հաշվարկում կատարել պահպանելով մշակման ալգորիթմի հատկությունները: Վերջին տարիների ընթացքում ներդրոնային ցանցերը նման պատկերների մշակման խնդիրներում ցուցաբերել են արդյունավետության զգալի բարձրացում:

Արդյունավետությունը ավելի բարձրացնելու համար հեղինակը CPU-GPU համակարգերի համար օգտագործել է խորը ուսուցման մոդելների JIT կոմպիլյատոր և հաշվարկման բաշխված համակարգ ամպային ծառայություններում հաշվարկներ կատարելու համար: Այդ համակարգը պարունակում է բաշխված զանգվածների պահպանման, վիրտուալ մեքենաների նախապատրաստման, կոնտեյներների տեղակայման և ցանցերի կառավարման տարրեր: Այդ մշակման թեստավորումը կատարվել է 10,000 CPU միջուկի և 300 GPU սարքերի վրա հասնելով 30 TFLOP հաշվարկման արտադրողականության: Համարժեք կենտրոնական պրոցեսորների վրա ներկայացված կոմպիլյատորը գերազանցում է լայն կիրառելի Tensorflow և PyTorch խորը ուսուցման միջավայրերին 3-4 անգամ և խնայում է 1.5 անգամ ռեսուրսների ծախսը:

Երբ էլեկտրոնային մանրադիտակը ձեռք է բերում կենսաբժշկական պատկերները շերտ առ շերտ, նրանք ընթացքում կարող են դեֆորմացվել կամ վնասվել: Հեղինակը դիտարկում է կենսաբժշկական պատկերների վերադրման և ծավալային արտապատկերման խնդիրը: Աշխատանքում ներկայացվում է նորմալիզացված կորելացիայի արդյունքում ստացված նեյրոնային շերտը: Ապա, կիրառվում է մետրիկ ուսուցման մեթոդը սիամական ցանցի վրա որպեսզի կատարելագործի հաջորդական իզոտոպային պատկերները վերադրումը:

Մեծ ծավալի կենսաբժշկական պատկերները վերադրելու նպատակով խորը ուսուցման մոդելը վերոնշյալ ներկայացված համակարգի շնորհիվ կոմպիլացվում է և կատարում տարաբաշխված հաշարկ: Արդյունքում վերադրման սխալմունքը բարելավում է 2 մինչև 7 անգամ նախկին մեթոդների համեմատ:

Ատենախոսությունը բաղկացած է ներածությունից, չորս գլուխներից, եզրակացությունից և օգտագործված գրականության ցանկից:

Գլուխ 1-ը ծառայում է որպես ակնարկ, որտեղ ամփոփվում են ատենախոսության համար անհրաժեշտ մեքենայական ուսուցման ենթաոլորտները: Համապատասխանորեն՝

Ենթագլուխ 1.1-ում հետազոտվել և ամփոփվել են մեքենայական ուսուցման միջավայրերը, մեծածավալ տվյալների պահպանման մեթոդները, ամպային և տարաբաշխված համակարգերը:

Ենթագլուխ 1.2-ում ներկայացված են օպտիմիզացիայի, մեքենայական ուսուցման և խորը ուսուցման հիմունքները: Քննարկվել են խորը ուսուցման տարբեր մոտեցումները:

Ենթագլուխ 1.3-ում ներկայացվել է կենսաբժշկական պատկերների աշխատանքի յուրահատկությունները այդ թվում վերադրման խնդիրը և գրականությունում հայտնի մեթոդները:

Ցավոք առաջին գլխում բացակայում է բուն խնդրի/ենթախնդիրների անհրաժեշտության հիմնավորումը և դրա ձևակերպումը, ինչպես, օրինակ, կատարված է սույն կարծիքի սկզբում, ինչ բավականին դժվարեցնում է թեզի հետագա ընթերցումը:

Գլուխ 2-ը նվիրված է խորը ուսուցման տարաբաշխված հաշվարկման խնդրի ամպային լուծումներին մեծ տվյալների մշակման ժամանակ:

Ենթագլուխ 2.1-ում հեղինակը ուսումնասիրել է խորը ուսուցման տարաբաշխված հաշվարկման խնդիրը և արդյունքում առաջարկել մեծ տվյալների մշակման նպատակով համակարգ ամպային լուծումներ ստանալու համար:

Ենթագլուխ 2.2-ում ուսումնասիրվել են համակարգի հաշվողական մոդելը, ինտերֆեյսը և փոփոխականների հայտարարումը:

Ենթագլուխ 2.3-ում ներկայացվել է այդ թվում ֆայլային համակարգի, վիրտուալ մեքենաների նախապատրաստման, կոնտեյներների տեղակայման և ցանցերի կառավարման մանրամասները:

Ենթագլուխ 2.4-ում քննարկվել է համակարգի առանձնահատկությունները:

Աշխատանքի **երրորդ գլխում** նկարագրվում է լուծումներ կենսաբժշկական մեծ պատկերների առանձնահատկությունները հաշվի առնելու համար:

Ենթագլուխ 3.1-ում ստեղծված համակարգը թեստավորվել է մեքենայական ուսուցման տարբեր խնդիրներում և ցուցադրել մաքսիմալ 30 ՊՖԼՕՊ հաշվարկման արտադրողականությունը:

Ենթագլուխ 3.2-ում հեղինակը ներկայացրել է CPU համակարգերի համար խորը ուսուցման մոդելների յոՒ կոմպիլատոր: Այն կիրառում է մեթոդներ ինչպիսիք են խորը ուսուցման շերտերի միաձուլում, կենտրոնական պրոցեսորի հիշողության օպտիմալացում:

Արդյունքում համարժեք կենտրոնական պրոցեսորների վրա 3-4 անգամ գերազանցվում են լայն կիրառելի Tensorflow և PyTorch խորը ուսուցման միջավայրերը արտադրողականության տեսակետից: Բացի դրանից արդյունքում հաշվարկումները կատարվում են 1.5 անգամ ավելի էժան ամպային լուծումներում օգտագործված GPU-ների քանակի տեսակետից:

Ենթագլուխ 3.3-ում հեղինակը ներկայացնում է կենսաբժշկական խնդիրների վերադրման խորը ուսուցման մոդել: Արդյունքում այն լավացնում է նկարների վերադրման սխալմունքը 2-7 անգամ:

Ենթագլուխ 3.4-ում ուսումնասիրվել է վերադրման խնդիրը կիրառելով ծայրից ծայր խորը ուսուցման մոդելներ: Արդյունքում խորը ուսուցման մոդելը վերածվում է էլաստիկ առածգական մոդելի:

Չորրորդ գլուխում նկարագրվում են համակարգի իրականացման մանրամասները:

Ենթագլուխ 4.1 դիտարկվում է համակարգի սխալին հանդուրժողականությունը:

Ենթագլուխ 4.2-ում թեստավորվում է տվյալների պահպանման մոդելների ուսուցանումը հեռակա տվյալների հոսքի հիման վրա: Արդյունքում ստացվում են տեղական պահպանման համարժեք արդյունքներ, սակայն տվյալները կարող են մեծածավալ լինել:

Ենթագլուխ 4.3-ը նվիրված է նորմալիզացված փաթույթային շերտի իրագործման մանրասններին tensorflow գրադարանի դեպքում:

Եթե դիտարկել խնդրի ներկայիս աղբյուր դրվածքը, ապա առաջարկված լուծումների բազմության ամփոփ և արդյունավետ լինելու հետ կապված բազմաթիվ հարցեր են ծագում, օրինակ, որն է այդ տիրույթը, որտեղ առաջարկված մեթոդները ամենաարդյունավետն են, ինչու են ընտրված այս և ոչ ուրիշ հիմքային մեթոդները, ինչու է ընտրվել այս մոդելային խնդիրը և այլն:

Միևնույն ժամանակ, եթե դիտարկել սույն կարծիքում առաջարկված դրվածքը, շեշտը դնելով միջավայրի/հարթակի վրա, ապա կարելի ասել, որ ստացված արդյունքները, բացի նոր լինելուց, լիակատար պատասխանում են հետազոտման լրիվության և ամփոփ լինելու բոլոր հարցերին, իսկ վերը նշված հարցերը դառնում են պարզապես ուղենիշներ համակարգի զարգացման

ճանապարհային քարտեզի համար: Դրանք հնարավորություն են տալիս եզրակացնել, որ ատենախոսության մեջ առաջարկված է էական նորույթ պարունակող արդյունավետ մոտեցում ու ամփոփ լուծում արդիական և կիրառական կարևոր նշանակություն ունեցող խնդրի համար, որի բնականոն զարգացումը կարող է բերել նոր հետաքրքիր արդյունքների ու կիրառությունների:

Նշված դիտողությունը նվազեցնում է կատարված աշխատանքի արժեքը: Միևնույն ժամանակ աշխատանքն իսկապես արդիական է, պարունակում է կարևոր գործնական խնդրի ամփոփ ինքնատիպ լուծում և ունի բազմաթիվ կարևոր կիրառություններ ու զարգացման հեռանկարներ՝ նոր ուսուցառման մեթոդներ մշակելու համար:

Հիմնական արդյունքները տպագրված են, սեղմագրի բովանդակությունը լիովին համապատասխանում է ատենախոսությանը:

Ամփոփելով, Դավիթ Բունիայանի ատենախոսությունը լիովին համապատասխանում է Ե.13.04 մասնագիտության գծով ներկայացվող թեկնածուական ատենախոսությունների պահանջներին, իսկ հեղինակը անկասկած արժանի է տեխնիկական գիտությունների թեկնածուի աստիճանի շնորհմանը:

Պաշտոնական ընդդիմախոս՝
ՀՀ ԳԱԱ ակադեմիկոս,
Ֆիզ.մաթ.գիտ.դոկտոր, պրոֆեսոր

Ս.Կ. Շուքուրյան
30.03.2020

ԵՊՀ ՏՏ կրթական և հետազոտական կենտրոնի գիտական ղեկավար ՀՀ ԳԱԱ ակադեմիկոս, ֆիզիկամաթեմատիկական գիտությունների դոկտոր, պրոֆեսոր Սամվել Կիմի Շուքուրյանի ձեռքի ստորագրությունը հավաստում եմ՝

ԵՊՀ գիտական քարտուղար՝
պ.գ.թ.



Լ.Ս.Հովսեփյան