НасіровЕмілМехдієвичаспіранткафедриматематичноїінформатикиКиївськийнаціональнийуніверситетіменіТарасаШевченкаНазвадисертаціїПаралелізаціяневідємноїфакторизаціїрозрідженихлінгвістичнихматрицьтатензорівнадвеликоїрозмірностіШифртаназваспеціальностітеоретичніосновиінформатикитакібернетикиСпецрадаДКиївськогонаціональногоуніверситетуіменіТарасаШевченка

КиївськийнаціональнийуніверситетіменіТарасаШевченка

МіністерствоосвітиінаукиУкраїни

Кваліфікаційнанаукова

працянаправахрукопису

НАСІРОВЕМІЛМЕХДІЄВИЧ

УДК

ДИСЕРТАЦІЯ

ПАРАЛЕЛІЗАЦІЯНЕВІД’ЄМНОЇФАКТОРИЗАЦІЇРОЗРІДЖЕНИХ

ЛІНГВІСТИЧНИХМАТРИЦЬТАТЕНЗОРІВНАДВЕЛИКОЇРОЗМІРНОСТІ

–теоретичніосновиінформатикитакібернетики

Подаєтьсяназдобуттянауковогоступенякандидатафізикоматематичнихнаук

ДисертаціяміститьрезультативласнихдослідженьВикористанняідей

результатівітекстівіншихавторівмаютьпосиланнянавідповіднеджерело

ЕМНасіров

Науковийкерівник

МарченкоОлександрОлександрович

докторфізикоматематичнихнаукпрофесор

Київ–

Зміст

Умовніпозначення

Вступ

РозділАналізсучасногостанугалузіфакторизаціїматрицьтатензорів

надвеликоїрозмірності

Постановказадачіфакторизаціїрозрідженихматрицьтатензорів

надвеликоїрозмірності

Підходидоневід’ємноїфакторизаціїматрицьтатензорів

Лінгвістичніматрицітатензори

ВисновкидоРозділу

РозділПаралелізаціяневід’ємноїфакторизаціїрозрідженихматриць

надвеликоїрозмірності

Алгоритмневід’ємноїфакторизаціїматриць

Паралелізаціяалгоритмузвикористанням

Розподіленаверсіяалгоритму

Оптимізаціязберіганнятапередачірозрідженихматриць

Нормалізаціяненульовихелементівблоківматриці

Тестуванняшвидкодії

ВисновкидоРозділу

РозділПаралелізаціяневід’ємноїфакторизаціїрозрідженихтензорів

надвеликоїрозмірності

Невід’ємнафакторизаціятензору

Розподіленняалгоритмуневід’ємноїфакторизаціїтензорів



Зведеннязадачіневід’ємноїфакторизаціхтензоріврангудо

невід’ємноїфакторизаціїматриць

ВисновкидоРозділу

РозділБлочнодіагональнийпідхіддоневід’ємноїфакторизації

матрицьтатензорів

Аналізблочнодіагональногопідходу

Приведенняматрицідоблочнодіагональноїформи

Проблемаприведеннярозрідженихматрицьтатензорівдоблочнодіагональноговиду

ВисновкидоРозділу

РозділФормуваннятематичнихдіагональнихблоків

Тематичнемоделюваннятекстовогокорпусу

ВикористаннялатентногорозподілуДіріхледляприведення

матрицьтатензорівдоблочнодіагональногоформи

Підтримказбагаченнямоделіновимиданими

Порівнянняшвидкодії

ВисновкидоРозділу

Висновки

Списоквикористанихджерел

ДОДАТОКА

ДОДАТОКБ

ДОДАТОКВ



Умовніпозначення

–невід’ємнафакторизація

невід’ємнаматричнафакторизація

невід’ємнатензорнафакторизація

матриця–матрицязв’язностіСловоДокументщоміститьчастоту

входженнясловавдокументі

–об’ємпам’ятінеобхіднийдлязберіганняматриціХ

–кількістьненульовихелементівматриці

СОО–координатнийформатзберіганняматриць

форматстиснутихрозрідженихрядківзберіганняматриць

Сформатстиснутихрозрідженихстовпчиківзберіганняматриць

графічнийпроцесор

центральнийпроцесор

потоковиймультипроцесорграфічногопроцесору

потоковийпроцесорграфічногопроцесору

–виконанняоднієїкомандинадекількохпотоках

одиницявимірюванняшвидкодіїобчислювальнихприладів

Обчисленнязагальногопризначеннянаграфічнихпроцесорах

–операціїчитаннязаписуданих

Імовірніснийлатентний

семантичнийаналіз

–блочнодіагональнаформа



Вступ

АктуальністьтемидослідженняСьогоднідлябагатьохзадач

інформаційноїгалузіпостаєпроблемаавтоматичноговиділенняважливої

інформаціїзтекстутаздобуваннялінгвістичнихзнаньзвеликихнаборівтекстів

природноюмовоюОднієюзнайважливішихтанайскладнішихзадачє

розумінняприродномовноготекстуНевід’ємнаматричнататензорна

факторизаціявикористовуєтьсявкомп’ютернійлінгвістицідлярозв’язання

такихприкладнихзадачяккласифікаціякластеризаціятекстівтатермінів

побудовамірсемантичноїблизькостіавтоматичневиділеннялінгвістичних

структуртавідношеньдляавтоматизаціїстворенняописулінгвістичних

структуртаіншихврамкахметодулатентносемантичногоаналізуТакийпідхід

даєможливістьотриманняключовихознактекступриродноюмовоюякідають

уявленняпройогозмісттасмисласаметемиконцепціїтасемантика

Невід’ємнафакторизаціявеликихбагатовимірнихмасивівданихзнайшла

своєвикористанняівіншихгалузяхтаєшироковживаноютехнологієюдля

обробкизвуковихспектрограмабомязовоїактивностіобробкитааналізу

графічнихзображеньобробкивідеозображеньтабагатьохіншихзадачобробки

тааналізувеликихмасивівданих

Щевсерединіхфінськоюгрупоюдослідниківбулипроведені

дослідженняневідємногоматричногорозкладупідназвоюпозитивне

розкладанняматриціМетодставбільшшироковідомийякневідємний

матричнийрозкладпіслятогоякЛііСонгпобудувалиітеративнийалгоритмта

досліджувалийоговластивостіЗастосуванняневід’ємноїматричноїтатензорної

факторизаціїдляобробкиприродноїмовитапобудовионтологічнихбаззнань

досліджувалосьАнісімовимАВМарченкоООалебулиобмеженіобробкою

невеликихкорпусівданих

Обчисленняневід’ємноїфакторизаціїрозрідженихматрицьнадвеликої

розмірностіпредставляєособливийінтерестаактуальністьзточкизоруїї



застосуванняувеликихсистемахобробкиприродноїмовизагального

тематичногопризначеннянеобмеженихвикористаннямлишедлявузьких

предметнихобластейПрипобудовітакихсистемнеобхіднаобробканадвеликих

невід’ємнихматрицьтатензорівтакихякматриціСловаСтаттіаболінгвістичні

тензориПідметПрисудокДодатокДляпідтриманняякостіроботитакихсистем

вумовахстрімкогоростутекстовихданихтакожєпотребавпостійнійобробці

новихтекстівтадонаповненнясистеминовимиданимищовіснуючих

алгоритмахпотребуєбагаточасутаобчислювальнихресурсівдляповторної

факторизаціївсієїматриціаботензору

Використаннястандартнихалгоритмівфакторизаціїстаєпрактично

неможливимчерезполіноміальнускладністьіснуючихалгоритмівтаповнотуданоїзадачіРазомізнадвеликимирозмірамилінгвістичнихматрицьта

тензорівцепризводитьдонеприйнятновеликихоб’ємівобчисленьТакі

обчисленнянеможутьбутивиконанізареальнийчасбезефективного

розпаралеленнятавідповіднихобчислювальнихпотужностейДляобробки

великихкорпусівтекстівможебутивикористаневиконанняпаралельних

розрахунківнаграфічнихпроцесорахЦевсвоючергувимагаєрозробки

спеціальногометодупаралелізаціїматричнихтатензорнихобчисленьта

приведеннядоблочнодіагональноговиду

Сказаневищеобґрунтовуєтеоретичнутапрактичнуактуальність

дисертаційноїроботи

Звязокроботизнауковимипрограмамипланамитемами

Дисертаційнароботаєскладовоючастиноюнауковихробітяківелисяна

кафедріматематичноїінформатикифакультетукібернетикиКиївського

національногоуніверситетуіменіТарасаШевченкапривиконанні

фундаментальноїтеми“Створеннятеоретичнихосновметодівтапрограмних

засобівінтелектуалізаціїінформаційно–комунікаційнихтатрансформерних

технологій”державнийномерреєстрації–ррта

держбюджетноїнауководослідноїроботиРОЗРОБКАЛОГІКО



АЛГОРИТМІЧНИХМЕТОДІВДОСЛІДЖЕННЯФОРМАЛЬНИХМОДЕЛЕЙ

ПРИРОДНИХМОВТЗНДР№БФрр

МетаізадачідисертаційногодослідженняМетоюдисертаційноїроботи

єпобудовамоделейпаралельнихалгоритмівневід’ємноїфакторизації

лінгвістичнихрозрідженихматрицьтатензорівнадвеликоїрозмірностідля

зменшеннянеобхідногочасутаобчислювальнихресурсів

Зоглядунаметувроботіставлятьсятакізадачі

Розробитипаралельнийалгоритмфакторизаціїматрицьтатензорівдля

виконанняобчисленьнаграфічнихпроцесорах

Побудуватирозподіленийалгоритмфакторизаціїматрицьтатензорів

надвеликоїрозмірностідлявиконанняобчисленьнемережі

обчислювальнихвузлівтазменшеннянеобхідногочасу

Розробитиметодфакторизаціїлінгвістичнихматрицьтатензорівза

допомогоюприведеннядоблочнодіагональноговидудлязменшення

оброблюванихоб’ємівданих

Розробитиалгоритмформуваннятематичнихдіагональнихблоків

лінгвістичнихматрицьтатензорівзвикористаннямгенеративних

статистичнихмоделейдляотриманнядодатковихсемантичнихзв’язків

межтермами

Провестиекспериментизоцінкишвидкодіїзапропонованогоалгоритму

тапорівнятийогорезультатизіншимиалгоритмами

Об’єктдослідження–алгоритминевід’ємноїфакторизаціїматрицьта

тензорів

Предметдослідження–моделіпаралелізаціїалгоритмівневід’ємної

факторизаціїрозрідженихматрицьтатензорівнадвеликоїрозмірності

МетодидослідженняДослідженнябазуютьсянаметодахтаалгоритмах

багатовимірногоаналізуталінійноїалгебритензорногочисленняпаралельних



обчисленьтеоріїсинтаксичногоаналізуштучногоінтелектуметодики

побудовикомпютернолінгвістичнихсистем

НауковановизнаодержанихрезультатівУдисертаційнійроботі

розробленоновімоделіпаралелізаціїдлявирішеннязадачіфакторизації

надвеликихрозрідженихматрицьтатензорівіотриманотакіновінаукові

результати

Розробленоновумодельпаралелізаціїалгоритмівфакторизації

матрицьтатензорівзвиконаннямобчисленьнаграфічних

процесорахзоптимальнимпорядкомвиконаннямопераціїдля

зменшеннячасучитаннятазаписуданихвпам’ятьграфічних

процесорів

Розробленоновумодельрозподіленогоалгоритмуфакторизації

матрицьтатензорівнадвеликоїрозмірностіоптимізувавши

використанняммережевихресурсівдлязменшеннячасу

необхідногодляпередачіданих

Розробленоновийметодфакторизаціїрозрідженихлінгвістичних

матрицьтатензорівзадопомогоюприведеннядоблочнодіагональноговидузможливістюдонаповненняновимиданими

Впершерозробленоалгоритмформуваннятематичнихдіагональних

блоківлінгвістичнихматрицьтатензорівзвикористанняммоделі

латентногорозподілуДіріхле

Проведенетестуванняпоказалощореалізаціяданихалгоритмів

підвищуєшвидкодіюфакторизаціїматрицьтатензорівзменшує

необхідніобчислювальніресурситаоб’ємнеобхіднихоперацій

читаннязаписутамережевихоперацій

ТеоретичнеіпрактичнезначенняодержанихрезультатівНаукове

значенняроботиполягаєврозробцімоделіпаралелізаціїалгоритмів



факторизаціїлінгвістичнихматрицьтатензорівнадвеликоїрозмірності

побудованихзприродньомовнихтекстів

Практичнезначенняроботиполягаєвзначномупокращеннішвидкодії

алгоритмівфакторизаціїЗменшеннянеобхіднихобчислювальнихресурсівдля

виконанняневід’ємноїфакторизаціїнадвеликихматрицьтатензорівдозволяє

проводитиподальшідослідженнявумовахобмеженихресурсів

Отриманірезультативпроваджуютьсядлядослідженьуобластірозробки

засобівінтелектуальноїобробкитекстівприродноюмовоютавякостіматеріалів

длякурсівШтучнийінтелекттаКомпютерналінгвістиканафакультеті

компютернихнауктакібернетикиКиївськогонаціональногоуніверситетуімені

ТарасаШевченка

ОсобистийвнесокздобувачаВсірезультатидисертаційноїроботи

отриманіавторомсамостійносформульованіувиглядіалгоритмівтаметодівта

обґрунтованізпосиланняминавикористаніджерела

Зарезультатамидисертаціїопублікованоодинадцятьробітунаукових

фаховихвиданняхУкраїни–однастаттяунауковомужурналівнесеному

доміжнароднихнаукометричнихбаз

Уроботахопублікованихуспівавторстві

–устаттяхпошукачуналежатьрезультатироботинадрозробкою

моделейпаралелізаціїфакторизаціїматрицьтатензорівтапрограмнареалізація

розробленихмоделейпаралелізаціїалгоритмів

–утезахміжнароднихконференційпошукачуналежатьрезультати

роботинадрозробкоюпаралельнихтарозподіленихалгоритмівтаїхреалізації

проведенняекспериментівтапроведеннятестування

АпробаціярезультатівдисертаціїРезультатидисертаціїбули

представленінаміжкафедральнихсемінарахфакультетукомпютернихнаукта



кібернетикиКиївськогонаціональногоуніверситетуіменіТарасаШевченката

доповідалисьнаміжнароднихконференціяхзокремана

’

–

’

–

’Міжнароднанауковаконференціястудентіваспірантівта

молодихвченихТеоретичнітаприкладніаспектикібернетики

листопадароку

Іміжнароднанауковопрактичнаконференція

Математичнетапрограмнезабезпеченняінтелектуальнихсистемм

ДніпропетровськУкраїналистопадароку

’МіжнародноїнауковотехнічноїконференціїШТУЧНИЙ

ІНТЕЛЕКТТАІНТЕЛЕКТУАЛЬНІСИСТЕМИмБердянськУкраїна

–вересняр





ПублікаціїЗарезультатамидисертаціїопублікованонауковихпрацьу

томучислі–статтіуфаховихвиданняхнауковихпрацьзатвердженихМОН

Українистаттяужурналівнесеномудоміжнароднихнаукометричнихбаз

–публікаціїуматеріалахітезахнауковихконференцій

СтруктуратаобсягдисертаціїДисертаційнароботаскладаєтьсяіз

вступуп’ятирозділіввисновківспискувикористанихджерелтатрьох

додатківЗагальнийобсягроботистановитьсторінокосновнийтекстроботи

викладенонасторінкахсписоквикористанихджерелналічує

найменуваньнасторінкахТекстроботинаписанийукраїнськоюмовою

Висновки

Удисертаціїприділенобагатоувагипопереднімроботамтапроведено

ретельнийаналізіснуючихпідходівдоневід’ємноїфакторизаціїматрицьта

тензорівзточкизорушвидкодіїтаможливостівикористаннядляобробки

надвеликихлінгвістичнихматрицьтатензорівПротеіснуючіпідходимають

суттєвінедолікищоробитьнеможливимїхвикористанняввеликихсистемах

обробкиприродньоїмовиЦимзумовлюєтьсяважливістьпроведеного

дослідженняДетальнийоглядлітературиджерелтаподальшедослідженняі

розробкановихметодівтаалгоритмівдозволилозначнозменшитичас

факторизаціїтаоб’єминеобхіднихобчислювальнихресурсів

Удисертаційнійроботірозробленоновімоделіпаралелізаціїалгоритмів

факторизаціїнадвеликихрозрідженихлінгвістичнихматрицьтатензорів

локальнамодельпаралелізаціїзвикористаннямжорсткогодискудлязберігання

данихтапаралельнихобчисленьнаграфічнихпроцесорахірозподілений

алгоритмзвикористанняммережівузлівдлярозрахунківтавиконанням

паралельнихобчисленьнаграфічнихпроцесорахщодалозмогузменшити

необхіднийчасобчислювальніресурситаотриманаможливістьзбагачення

системобробкиприродньоїмовиновимиданимибезнеобхідностіповторної

факторизації

Врезультатіпроведеноїроботиотриманотакінауковірезультати

Длявиконанняобчисленьнаграфічнихпроцесорахзоптимальним

порядкомвиконаннямопераціїдлязменшеннячасучитаннятазапису

данихвпам’ятьграфічнихпроцесоріврозробленомодельпаралелізації

алгоритмівфакторизаціїматрицьтатензорів

Оптимізовановикористанняммережевихресурсівдлязменшеннячасу

необхідногодляпередачіданихрозробившимодельрозподіленого

алгоритмуфакторизаціїматрицьтатензорівнадвеликоїрозмірності



Розробленоновийметодфакторизаціїрозрідженихлінгвістичних

матрицьтатензорівзадопомогоюприведеннядоблочнодіагонального

видущодозволяєзменшитиоб’ємданихдляфакторизаціїтанадає

можливістьзбагаченнясистеминовимиданими

Впершерозробленоалгоритмформуваннятематичнихдіагональних

блоківлінгвістичнихматрицьтатензорівзвикористанняммоделі

латентногорозподілуДіріхлещодозволяєотриматинеобхіднукількість

діагональнихблоківатакожможливістьаналізудодатковихзв’язківміж

блокамитатермами

Проведенотестуваннятапорівнянняшвидкодіїрозробленихалгоритмів

таметодівнанадвеликійматрицізв’язностіСловаСтаттітатензорі

ПідметПрисудокДодатокЕкспериментипоказалищореалізація

запропонованихалгоритмівпідвищуєшвидкодіюфакторизаціїматриць

татензорівзменшуєнеобхідніобчислювальніресурситаоб’єм

необхіднихопераційчитаннязаписутамережевихоперацій