

НЕЙРОН ТАРМОҚНИ ЎҚИТИШНИ ТЕЗЛАШТИРИШДА MKL
ПАРАЛЛЕЛЛАШТИРИШ КУТУБХОНАСИНING ИМКОНИАТИ

Очиллов Маннон Муслимович¹, Жавлиев Шахзод Алишер ўғли²

¹ Муҳаммад ал-Хорзамий номидаги Тошкент ахборот технологиялари университети,
Сунъий интеллект кафедраси доценти, PhD.

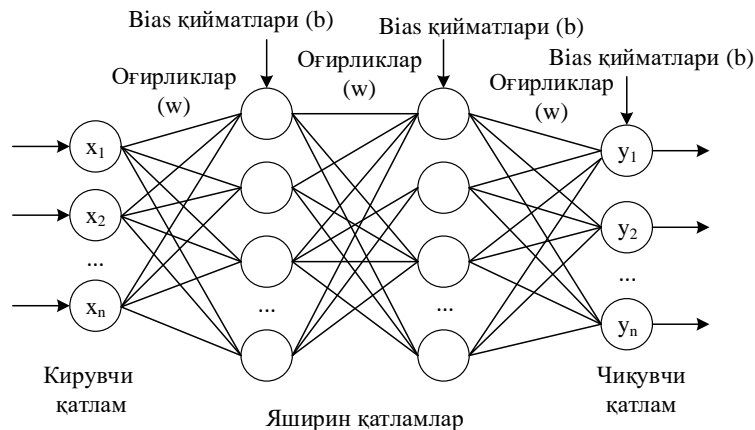
² Муҳаммад ал-Хорзамий номидаги Тошкент ахборот технологиялари университети
стажёр тадқиқотчиси

<https://doi.org/10.5281/zenodo.7856100>

Abstract. *In this work, tools that allow saving time resources in the process of training a neural network based on artificial intelligence and information about their application are given. Here we consider how to speed up neural network training using the Intel Math Kernel Library (MKL) and present the comparative results obtained during the research.*

Keywords: *Artificial Intelligence, Machine Learning, Neural Network, CPU, GPU, Parallel Processing, MKL Parallelization Library*

Технологияларнинг ривожланиши натижасида компьютерларнинг ишлаш тезлиги кундан-кунга тезлашиб бормоқда. Ривожланиб бораётган сунъий интеллект соҳасида ҳам машинали ўқитиш жараёнларини амалга оширишда кўп вақт талаб этиб бормоқда. Сўнги пайтларда ҳар қандай соҳани ахборот технологияларисиз, сунъий интеллектсиз тасаввур қилиш қийин. Сунъий интеллект ва машинали ўқитиш маълумотлар билан боғлиқ бўлган деярли барча соҳада жадал ўсиб бораётганини гувоҳи бўлишимиз мумкин [3]. Машинали ўқитиш (МУ) – бу сунъий интеллектнинг бир тури бўлиб, компьютерни дастурлашнинг янги йўналиши ҳамда дастурнинг маълумотларни ўқиб олиш ва шу маълумотлар асосида ўрганиш хусусиятига эга бўлиши, компьютер маълумотларини ўқиб олиш асосида хусусиятларни ўрганиш қобилиятидир. Машинали ўқитиш ва нейрон тармоқлар бир-бири билан чамбарчас боғлиқ тушунчалардир. Нейрон тармоқлар (НТ) - бу инсон миясининг хатти-ҳаракатларига тақлид қилиш учун мўлжалланган машинали ўқитиш моделининг бир туридир. Нейрон тармоғида кириш маълумотлари қатламларга ташкил этилган бир қатор ўзаро боғланган тугунлар ёки нейронлар орқали қайта ишланади. Ҳар бир нейрон бошқа нейронлардан маълумот олади ва киритилган маълумотларнинг оғирлаштирилган йиғиндиси асосида чиқиш қийматини ҳосил қилади. Ушбу уланишларнинг оғирликлари ўқув жараёнида оптималлаштириш алгоритми, масалан, градиент тушиши ёрдамида ўрганилади. Нейрон тармоқлар тасвирларни таснифлаш, нутқни аниқлаш ва табиий тилни қайта ишлаш каби турли хил машинали ўқитиш вазифаларида ажойиб самарадорликни кўрсатди. Қуйидаги 1-расмда тўлиқ боғланишли нейрон тармоқ архитектураси келтирилган.

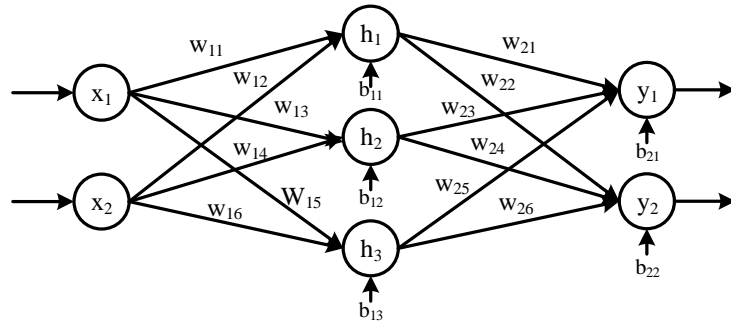


1-расм. Тўлиқ боғланишли нейрон тармоқ архитектураси

Нейрон тармоқларнинг асосий афзалликларидан бири бу одамлар томонидан аниқ лойиҳалашни талаб қилмасдан, маълумотлар тасвирларини автоматик равишда ўрганиш қобилиятидир. Бу самарали моделларни ишлаб чиқиш учун талаб қилинадиган соҳага оид билимлар миқдорини сезиларли даражада камайтирди, лекин нейрон тармоқни ўқитишда ҳамда мураккаб маълумотларни таҳлил қилишда узоқ вақт ёки машаққатли ҳисоблаш меҳнатини талаб қилади. Шу ўринда машинали ўқитишда тасвир, видео, матн, нутқ сигналлари, турли хил ўлчамдаги матрицалар устида ҳисоблаш ишларини олиб бориш кўп вақтни талаб этади.

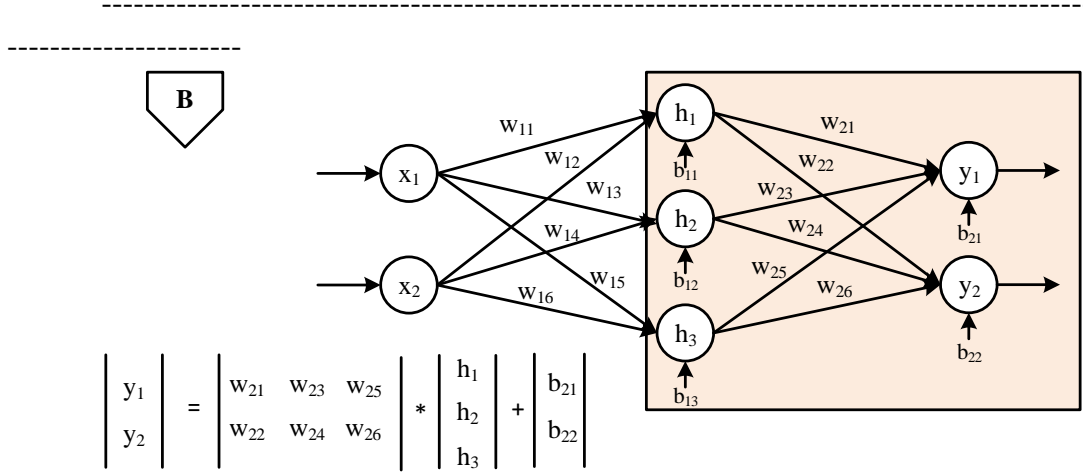
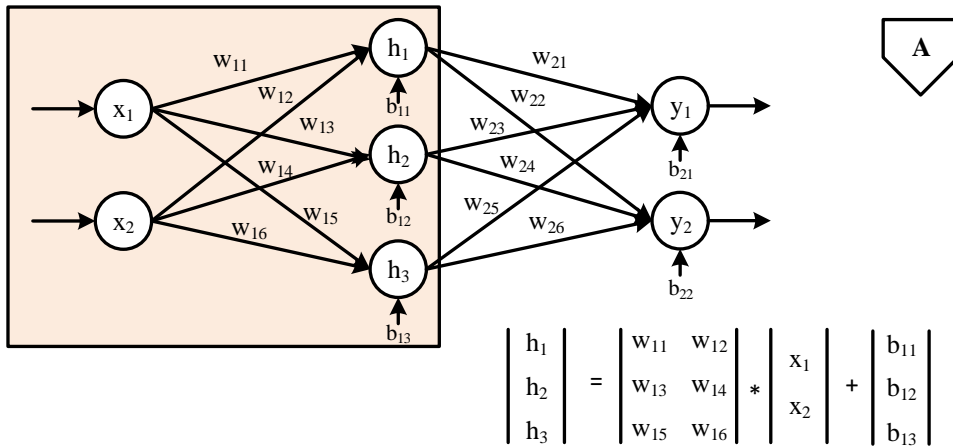
Ҳозирги вақтда турли соҳалар қаторида нейрон тармоқларни ўқитиш юқори унумдорликдаги параллел ҳисоблаш воситаларини талаб этмоқда [4,5]. Параллел ишлов бериш - бу компьютерда бир нечта процессорлар ёки ядроларда бир вақтнинг ўзида бажариладиган иш тартиби ҳисобланади. Одатда, жараёнлар ўртасида алоқа ўрнатишда бироз қўшимча юклама ҳосил бўлади. Бу эса кичик ҳажмдаги вазифалар учун ҳам умумий вақтни қисқартириш ўрнига аксинча ошиб кетишига олиб келиши мумкин. Бу муаммони Intel Parallel Studio таркибига кирувчи кутубхоналардан бири MKL (Intel Math Kernel Library) кутубхонаси орқали ҳал қилиш мумкин.

Intel Math Kernel Library (MKL) - бу Intel процессорларида нейрон тармоқ ҳисоблашларини тезлаштириш учун ишлатилиши мумкин бўлган оптималлаштирилган математик амаллар кутубхонасидир. MKL нейрон тармоқларни ўқитиш ва хулоса чиқаришни тезлаштириш учун ишлатилиши мумкин бўлган бир қатор параллел алгоритмларни ўз ичига олади. Кутубхона ўз ичига математик операциялари ва алгоритмларни оптималлаштирилган ҳолда амалга ошириш имконини яратади ва нейрон тармоғини ҳисоблашда тез-тез ишлатиладиган асосий чизиқли алгебра кичик дастурларнинг оптималлаштирилган иловаларини ўз ичига олади. Буларга матрицаларни кўпайтириш, векторларни кўшиш ва нейрон тармоқларни ҳисоблашда кенг қўлланиладиган бошқа чизиқли алгебра операциялари учун амаллар киради. Масалан кириш қатлами 2 та нейрондан, яширин қатлам 3 та нейрондан ва чикувчи қатлам эса 2 та нейрондан иборат куйидагича тармоқ берилган бўлсин (2-расм).



2-расм. Бир яширин қатламли нейрон тармоқ архитектураси

Бунда матрица устида амалга ошириладиган амаллар, кўшиш ва кўпайтириш амалларидан фойдаланган ҳолда кирувчи қатлам нейронлари мос равишда уларнинг оғирлик коэффициентларига (*weights*, w_{ij}) кўпайтирилади ва ҳар бир натижавий нейронга озод ҳад (*bias*, b_{ij}) қўшилади (3-расм А ва В чизмалар).



3-расм. А) Яширин қатлам қийматларини ҳисоблаш, В) Чикувчи қатлам қийматларини ҳисоблаш

Нейрон тармоқни ўқитиш ҳамда унда ҳисоблаш жараёнларини кўриб чиқдик. Эндиликда биз ушбу ўқитилган нейрон тармоқни ҳисоблаш жараёнларини амалга оширишда тезкорликка эришиш йўллари ва унинг натижаларини кўриб чиқамиз. Йўқорида таъкидлаб ўтилган нейрон тармоқни ҳисоблашда параллел ишлов бериш учун

**INTERNATIONAL SCIENTIFIC AND TECHNICAL CONFERENCE
“DIGITAL TECHNOLOGIES: PROBLEMS AND SOLUTIONS OF PRACTICAL
IMPLEMENTATION IN THE SPHERES”**

APRIL 27-28, 2023

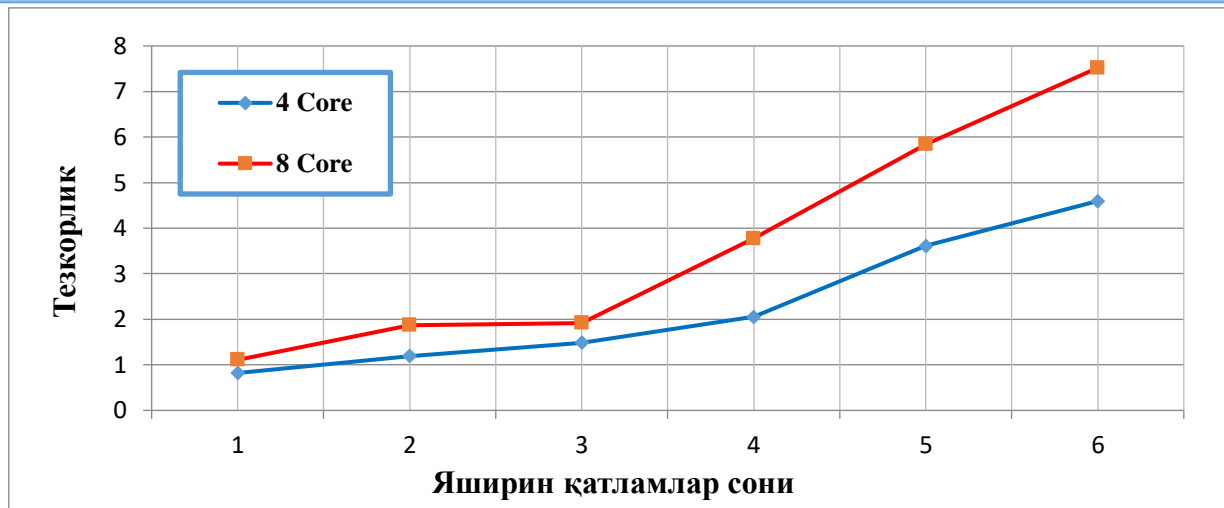
мўлжалланган кутубхоналардан бири Intel MKL кутубхонасидан фойдаландик. Юқорида таъкидлаб ўтилган MKL математик ҳисоблаш ишларини оптималлашган ҳолда амалга оширади. У векторлар ва матрицалар устида бажариладиган барча амалларни компьютер процессор ядроларида тақсимлаган ҳолда параллел ҳисоблайди ва эффе́ктелиги бошқа праллеллаштириш муҳитларидан бутунлай фарқ қилади [1]. Мисол учун икки матрицани кўпайтиришни кўриб ўтадиган бўлсак умумдорлик юқори кўрсаткичга эришганини яққол кўришимиз мумкин. MKL шунингдек, нейрон тармоқ учун оптималлаштирилган алгоритмларни ўз ичига олади. Бу кўплаб нейрон тармоқ архитектураларида, айниқса тасвир ва видеоларни қайта ишлашда ишлатиладиган нейрон тармоқларнинг асосий амалларини бажаришда қўлланилади. MKL нинг ишлаш алгоритмлари кўп ядроли процессорларда ҳисоблашни тезлаштириш учун оқим ва жараён даражасида параллелликдан фойдаланади [1,2]. MKL кутубхонаси чизиқли алгебра вазифлари ҳисобланган матрица ва векторлар устида амаллар бажаришда жуда юқори самара беради. Биз MKL кутубхонаси нейрон тармоқнинг ўқитишни амалга оширишдаги самарадорлигини баҳолаб кўрдик. Тажрибаларимизда кўлда ёзилган араб рақамларини синфлаштириш масаласи объект сифатида қаралди. Нейрон тармоқни ўқитишни MNIST датасетидан фойдаланиб амалга оширдик. Унга кўра MNIST датасетида 10 та синф ва 60000 та намуна бор бўлиб ҳар бир намуна 28x28 ўлчамда ифодаланган. Тажрибаларимизда кўриб ўтган нейрон тармоқда кирувчи қатламдаги нейронлар сони 784 та ва чикувчи қатламдаги нейронлар сони 10 та. Яширин қатламлардаги тугунлар сонини 512 этиб белгилаб олдик ва яширин қатламлар сони ўзгартирилгандаги MKL кутубхонасининг самарадорлигини баҳоладик. Ўқитиш қадамларини сони 50 та. Унумдорлик кўрсаткичларини мос равишда қуйида келтирилган 1-жадвал ва 5-расмда кўришимиз мумкин. Изоҳ: Ишда нейрон тармоқнинг аниқлик кўрсаткичига эътибор қаратилмаган ва асосий мақсад ўқитиш вақтини қисқартириш деб қаралган.

1-жадвал.

MKL параллеллаштириш муҳитида процессорлар унумдорлиги

Яширин қатламлар сони	1 Core	4 Core		8 Core	
	Вақт (сек)	Вақт (сек)	Тезкорлик (марта)	Вақт (сек)	Тезкорлик (марта)
1	123,2	150,1	0,821	111,1	1,109
2	167,3	140,5	1,191	89,4	1,872
3	210,8	142,0	1,485	138,6	1,921
4	233,2	113,6	2,053	61,7	3,777
5	261,5	72,4	3,611	54,0	5,843
6	338,4	73,6	4,596	45,0	7,520

Ушбу жадвалда келтирилган маълумотларга қараганда, компьютеримизнинг процессорлар сони ортиб борган сари нейрон тармоқни ўқитишда тезкорлигимиз йўқори самарадорликка эришишилганини кўришимиз мумкин. Шунингдек, нейрон тармоқда ҳисоблаш ҳажми қанча ошса параллеллаштириш кутубхонасининг самарадорлиги юқори бўлади. Нейрон тармоқ яширин қатламлари сони ва CPU ядролари сонига мос самарадорликнинг ўзгариши қуйидаги 5-расмда келтирилган.



5-расм. MKL параллеллаштириш кутубхонасининг самарадорлиги

Хулоса сифатида шуни таъкидлаб ўтиш жоизки, нейрон тармоқлар машинали ўқитиш соҳасидаги асосий восита бўлиб, маълумотлардан ўрганиш ва янги, кўринмас мисолларга умумлаштириш мумкин бўлган кучли моделларни ишлаб чиқишга имкон берди. MKL нинг оптималлаштирилган параллел алгоритмларидан фойдаланган ҳолда, ишлаб чиқувчилар Интел процессорларида нейрон тармоқда бўладиган ҳисоблашларини сезиларли даражада тезлаштириши мумкин. Бу айниқса, катта ҳажмдаги нейрон тармоқ моделлари учун муҳим бўлиши мумкин, уларни ўқитиш узок вақт талаб қилиши ва муҳим ҳисоблаш ресурсларини талаб қилиши мумкин.

REFERENCES

1. A. Kalinkin, A. Anders, and R. Anders. “Intel Math Kernel Library PARDISO for Intel Xeon Phi TM Manycore Coprocessor”. In: Applied Mathematics 6.08 (2015), p. 1276
2. Khujayarov Ilyos Shiraliyevich, Islomov Shahboz Zokir ugli, Nuriev Sirojiddin Abduraxmonovich, Ochilov Mannon Musinovich. “Parallel Solving Tasks of Digital Image Processing”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 4, no. 6, June 2016, pp. 335-9, doi:10.17762/ijritcc.v4i6.2318.
3. Mekhriddin Rakhimov, Ravshanjon Akhmadjonov, Shahzod Javliev, “Artificial Intelligence in Medicine for Chronic Disease Classification Using Machine Learning”, 2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT), 2022, pp. 1-6, DOI: 10.1109/AICT55583.2022.10013587.
4. M. Rakhimov, J. Elov, U. Khamdamov, S. Aminov and S. Javliev, "Parallel Implementation of Real-Time Object Detection using OpenMP," 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670146.
5. M. Musaev and M. Rakhimov, "Accelerated Training for Convolutional Neural Networks," 2020 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2020, pp. 1-5, doi: 10.1109/ICISCT50599.2020.9351371.