

Dərin Neyron Şəbəkələrin Çoxnövəli Qrafik Prosesor Platformasında Öyrədilməsi

Kamil Ayda-zadə¹, Elşən Mustafayev², Samir Rüstəmov³

^{1,2}AMEA İdarəetmə Sistemləri İnstitutu, Bakı, Azərbaycan

³ADA Universiteti, Bakı, Azərbaycan

¹kamil-aydazade@rambler.ru, ²elshan.mustafayev@gmail.com, ³samir.rustamov@gmail.com

Xülasə— Məqalədə dərin neyron şəbəkələrin çox prosessorlu qrafik GPU və ənənəvi CPU-əsaslı platformalarında öyrədilməsi məsələsinə baxılmışdır. Test məsələ olaraq MNIST (Modified National Institute of Standards and Technology) verilənlər bazasında əlyazma rəqəmlərinin tanınması götürülmüşdür. Dərin neyron şəbəkənin verilmiş məsələ üçün hər iki platformada nəticələri müqayisəli təhlil edilmişdir.

Açar sözlər— dərin neyron şəbəkə, rəqəmlərin tanınması, GPU, CUDA, Tensorflow, MNIST

I. GİRİŞ

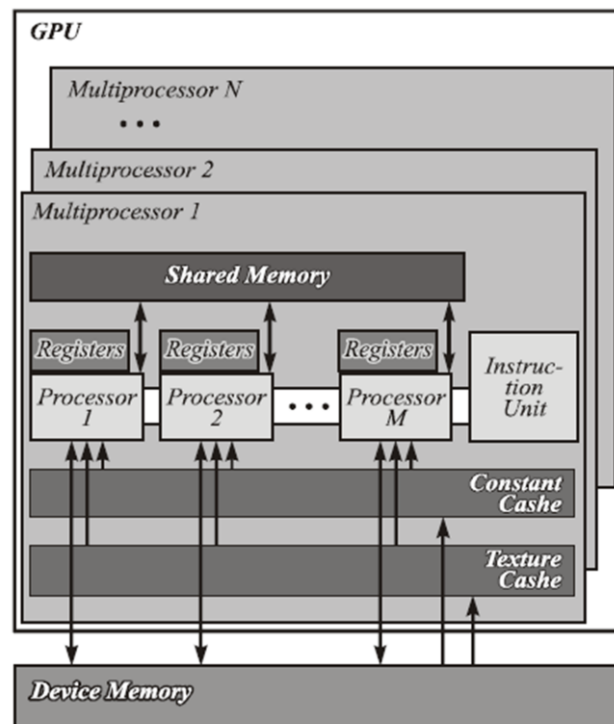
Son zamanlar sürətlərin tanınmasında (təsvirlərin klassifikasiyası və klasterizasiyası, səs və qrafik məlumatların tanınması və s.) dərin neyron şəbəkələrin istifadəsi geniş yayılıb. Bu növ şəbəkələrin qurulması və öyrədilməsi üçün müxtəlif proqram sistemləri işlənmişdir – Tensorflow, Caffe, Torch, Teano. Bu paketlərin köməyi ilə dərin neyron şəbəkələrin məsələlərini asanlıqla həll etmək olur. Lakin dərin neyron şəbəkələrin inkişafının ən böyük məhdudiyyəti yüksək hesablaşma xərcləridir. Bu məhdudiyyətlərin aradan qaldırılması yollarından biri kimi çoxnövəli qrafik prosessorlardan istifadə nəzərdə tutulur. Belə həllərin arasında NVIDIA şirkətinin CUDA (Compute Unified Device Architecture) texnologiyası xüsusi yer tutur. CUDA texnologiyası çoxnövəli qrafik prosessorların (GPU) nüvələrini paralel istifadə etməklə elmi və mühəndis hesablamalarını müasir klaster sistemləri səviyyəsində sürətləndirir [1].

Təqdim etdiyimiz məqalənin məqsədi Tensorflow [2] sistemində CUDA texnologiyasını istifadə etməklə əlyazma rəqəmlərinin dərin neyron şəbəkə vasitəsi ilə tanınması məsələsində imkan və xüsusiyyətlərin araşdırılmasıdır.

II. CUDA MÜHİTİNDƏ PARALEL HESABLAMALARIN APARILMASI

NVIDIA qrafik sürətləndiricilərində səkkizinci seriyadan başlayaraq CUDA paralel arxitekturası reallaşdırılıb. Bu texnologiya qeyri-qrafik hesablamaları aparmaq üçün xüsusi proqram interfeysini təqdim edir [3].

Məntiqi cəhətdən CUDA texnologiyasını dəstəkləyən qrafik prosessorlara çoxnövəli prosessorlar toplusu kimi baxmaq olar. Belə videoçiplərin əsas hesablaşma blokları multiprosessorlardan ibarətdir. Onlar səkkiz nüvədən, bir-neçə min 32-bitlik registrdən, 16 kбайt ümumi yaddaşdan, tekstur və konstant keşdən ibarətdir (şək.1).



Şəkil 1. Video kartın multiprosessor toplusu və yaddaş növləri.

CPU və GPU arasında əlaqə qurmaq üçün qlobal yaddaş istifadə olunur və məlumatların ötürülməsi sürəti yalnız PCI Express buraxma imkanı ilə məhdudlaşır. CUDA texnologiyasının proqramlaşdırma nöqtəyi-nəzərindən əsas mahiyyəti paralel olaraq eyni növ hesablamaların vahid iyerarxik hesablaşma axınına keçməsidir. Belə çoxsaylı hesablaşma axınları bir multi-prosessorada yerinə yetirilir və ümumi yaddaş vasitəsi ilə qalan multi-prosessorlarla sinxronlaşılır.

III. EKSPERİMENTLƏRİN NƏTİCƏLƏRİ

Qrafik prosessorların dərin neyron şəbəkələrində hesablaşma üstünlüklərini yoxlamaq üçün NVIDIA Tesla K40 qrafik sürətləndiricidən istifadə olunub. Bu video sürətləndiricinin göstəriciləri aşağıdakılardır:

- Maksimal hesablaşma gücü - 4.29 Tflops
- Yaddaş sürəti - 288 GB/s

- Yaddaş həcmi - 12GB
- Təcrübə zamanı istifadə olunan platformanın göstəriciləri:
- Prosessor – İntel Core i3-3.3GHz;
- Operativ yaddaş – 4GB;
- Qrafik sürətləndirici – NVIDIA Tesla K40;
- Proqram təminatı:
 - ✓ Windows 7 64-bit Professional;
 - ✓ Python 3.5.2;
 - ✓ Tensorflow 1 GPU support;
 - ✓ CUDA Toolkit 8.0.

Eksperimentlərdə əlyazma verilənlər bazası kimi məşhur MNIST [4] bazasından istifadə olunub. Bu verilənlər bazası 55000 öyrədici və 5000 test nümunələri olan əlyazma rəqəm təsvirləridir. Bütün təsvirlər ölçü baxımından normallaşdırılıb və 28*28 ölçüyə gətirilmişdir. Təsvirlərdəki piksellər 0-255 arasında qiymətlər alır. Burada 0-qara, 255-ağ rəngi göstərir, arada olan qiymətlər isə boz rəngin çalarlarını bildirir (şək. 2).



Şəkil 2. MNIST əlyazma nümunələri.

Təcrübədə iki müxtəlif strukturlu neyron şəbəkələrdən istifadə olunub:

1. Dörd laya malik tam əlaqələndirilmiş şəbəkə - [5]:
 - ✓ giriş layı 28x28=784
 - ✓ 1-3 gizli lay: 500 neyron
 - ✓ çıxış lay: 10 neyron
2. Bürünmə şəbəkəsi - *Convolutional Network* (CN) [6]:
 - ✓ giriş layı : 28x28=784
 - ✓ 1-ci gizli lay:
 - Bürünən lay: 1giriş, 32 çıxış təsviri, 5x5 ölçülü nüvə
 - MaxPool algoritmi
 - ✓ 2-ci gizli lay:
 - Bürünən lay: 32 giriş, 64 çıxış təsviri, 5x5 ölçülü nüvə

- MaxPool algoritmi
- ✓ 3-cü lay - tam əlaqələndirilmiş
 - giriş layı – 1024 neyron
 - çıxış layı - 10 neyron.

Öyrətmə üsulu olaraq Adam optimallaşdırılma üsulu istifadə edilmişdir.

Cədvəl 1-dən göründüyü kimi, qrafik prosessorlardan istifadə neyron şəbəkələrin öyrədilməsini əhəmiyyətli dərəcədə sürətləndir.

CƏDVƏL I. NEYRON ŞƏBƏKƏLƏRİN ÖYRƏTMƏ NƏTİCƏLƏRİNİN MÜQAYİSƏSİ

Neyron şəbəkə	Təcrübə platforması		Tanınma nəticələri	
	CPU	CPU + GPU	Öyrətmə baza	Test baza
MN	6 dəq 31 s.	2 dəq 25 s.	91,6%	89,7%
CN	47 dəq 5 s.	6 dəq 21 s.	99,6%	98,2%

ƏDƏBİYYAT

- [1] M. Fatica. CUDA for High Performance Computing: materials of HPC-NA Workshop 3, January, 2009
- [2] <https://www.tensorflow.org/>
- [3] CUDA Toolkit | NVIDIA Developer <https://developer.nvidia.com/cuda-toolkit>
- [4] Y. Le Cun. The MNIST database of handwritten digits // MNIST handwritten digit database, Yann LeCun and Corinna Cortes. [2009]. <http://yann.lecun.com/exdb/mnist>
- [5] К.Р.Айда-заде, Э.Э.Мустафаев. Об оптимизации параметров нейронной сети на этапе ее обучения / Труды Республиканской научной конференции «Современные проблемы информатизации, кибернетики и информационных технологий», ", том I, Баку, 2003, с. 118-121
- [6] Y.Le Cun, Y.Bengio. Convolutional networks for images, speech, and time-series. In M. A. Arbib, editor, The Handbook of Brain Theory and Neural Networks. MIT.

Bu iş Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı Fondunun maliyyə yardımı ilə yerinə yetirilmişdir - 5-ci mobillik qrant.