



Universiteit
Leiden

The Netherlands

Tuning the tuner: the art of automatic performance optimization

Willemsen, F.Q.

Citation

Willemsen, F. Q. (2026, April 14). *Tuning the tuner: the art of automatic performance optimization*. Retrieved from <https://hdl.handle.net/1887/4301430>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4301430>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

Moderne wetenschappelijke ontdekkingen en technologieën zijn in toenemende mate afhankelijk van krachtige computers. Van het voorspellen van het klimaat tot het trainen van kunstmatige intelligentie vertrouwen onderzoekers op High-Performance Computing (HPC)-systemen die traditionele processoren combineren met gespecialiseerde accelerators zoals grafische kaarten (GPU's). Deze machines zijn extreem snel, maar het is verrassend lastig om er het maximale uit te halen. Kleine keuzes in de manier waarop een programma wordt geschreven, zoals hoe data worden opgeslagen, hoe berekeningen worden georganiseerd, of hoeveel parallele verwerkingen er zijn, kunnen leiden tot grote verschillen in prestaties.

Omdat het aantal mogelijke keuzes enorm is, bestaat het automatisch verbeteren van prestaties (*auto-tuning*): de computer zelf automatisch verschillende configuraties van een programma laten testen totdat er een snelle en efficiënte variant wordt gevonden. Auto-tuning is zeer effectief gebleken, maar introduceert ook een nieuw probleem: hoe ontwerpen we dit afstemmingsproces zelf het beste? De keuze van de zoekstrategie, welke waarden getest worden en hoe om te gaan met ongeldige configuraties kan een groot verschil maken.

Dit proefschrift richt zich op het verbeteren van auto-tuning voor GPU's, die veel worden gebruikt in zowel wetenschappelijke als industriële toepassingen. Het onderzoek levert bijdragen in zes richtingen:

- Ontwikkeling van een nieuwe optimalisatiemethode gebaseerd op bayesiaanse statistiek, speciaal ontworpen voor GPU-tuning, die bestaande benaderingen overtreft.
- Het creëren van een eerlijke manier om tuningmethoden te vergelijken, zodat onderzoekers betrouwbaar kunnen beoordelen welke aanpak het beste werkt.
- Bouw van een snelle oplosser voor zoekruimtes, die miljarden mogelijke configuraties in minder dan een seconde kan verwerken en zo een belangrijke bottleneck wegneemt.

Samenvatting

- Aantonen dat het tunen van de tuners zelf kan leiden tot prestatieverbeteringen van meer dan 200%, terwijl het energieverbruik wordt verminderd via grootschalige simulaties.
- Evolutionaire algoritmen beter beperking-bewust maken, zodat zij effectiever kunnen omgaan met complexe beperkingen die inherent zijn aan GPU-programma's.
- Onderzoeken of kunstmatige intelligentie nieuwe optimalisatiealgoritmen kan vinden, en laten zien dat deze automatisch ontworpen algoritmes beter kunnen presteren dan die van mensen.

Gezamenlijk laten deze bijdragen zien dat het verbeteren van de onderliggende tuning-systemen zelf de volgende generatie technologische doorbraken kan voortbrengen. Alle methoden en hulpmiddelen uit dit proefschrift zijn als open-source software beschikbaar gesteld, zodat onderzoekers en praktijkmensen wereldwijd ervan kunnen profiteren.

Vooruitkijkend wijzen de resultaten op een toekomst waarin auto-tuning slimmer, efficiënter en adaptiever wordt. Algoritmen die beperkingen van de zoekruimte begrijpen, prestaties afwegen tegen energieverbruik, of zelfs zichzelf ontwerpen met behulp van AI, kunnen rekenkracht op grote schaal toegankelijker en duurzamer maken. Op die manier draagt dit werk bij aan vooruitgang in wetenschap en technologie, aangedreven door computers die zichzelf afstemmen op de uitdagingen van morgen.