Heungsub Lee

Contact
    heungsub@subl.ee or +82 10-3215-2380
Web Sites
    subl.ee, github.com/sublee, linkedin.com/in/sublee

Interests

- Back-end engineering for ML projects
- Developing and documenting API for engineers or researchers
- Free and open-source software

Skills

Programming Languages
    Python (expert), Go, Bash, C#, JavaScript
AI Research
    PyTorch, data or pipeline parallelism, Triton Inference Server, NVIDIA Nsight Systems
Back-end Development
    Linux, Docker, Kubernetes, AWS, Terraform, ZeroMQ, Redis, Couchbase, MySQL, etcd

Work Experience

Software Engineer
Naver, 2020–
A technology company.

Redesign the research platform to boost productivity in Clova.

Software Engineer
Kakao Brain, 2018–2020
The AI research lab in Kakao.

Focused on parallel training of convolutional neural networks, including large-scale data parallelism and pipeline parallelism. Developed and published a pipeline parallelism library in PyTorch named torchgpipe.

Developed a serverless training framework and a distributed hyperparameter search platform for an AutoML service.

Game Server Engineer & Architect
Nexon, 2011–2018
Developed and launched Durango, KartRider Dash & Coin Rush.

Designed and developed cloud-based distributed game servers for Durango (MMORPG) and KartRider Dash & Coin Rush (online racing games) respectively. Durango achieved up to 70k concurrent users per game world.

Developed an internationalization and localization system focused on linguistic features of Korean and Indo-European languages.

Researched rating systems such as Elo, Glicko, and TrueSkill to develop a matchmaker.

Led a server engineering team including 15 engineers.

Web Developer
Npine, 2008–2011
Supplies stock images for business on Iclickart.

Developed web services from scratch. Maintained on-premise Linux servers.

Front-end Web Developer
Lunant, 2008–2011
Served social media named VLAAH.

Designed and implemented the UI/UX for social media.

Open Source Experience

torchgpipe, 2019–2020
A GPipe implementation in PyTorch.

Implemented GPipe in PyTorch with optimization for CUDA, PyTorch's autograd engine, and long skip connections. GPipe is a scalable pipeline parallelism library for the training of a giant model.

This project has become a part of PyTorch. The story behind it can be found on Kakao Brain Blog[ko] and the technical report.

Hangulize, 2010–
Automatically transcribes a non-Korean word into Hangul.

Implemented an automatic Hangul transcription algorithm to realize Brian Jongseong Park's idea. Designed and implemented the web service and RESTful API.

Many professional Korean translators habitually use this tool to translate undocumented proper nouns. Netflix refers to this tool in the Korean timed text style guide.

TrueSkill, 2012–
A TrueSkill™ implementation in Python.

Implemented TrueSkill™, which is a rating algorithm for Xbox Live, in Python with a handy interface. This project was introduced in PyData Berlin 2019.

Profiling, 2014–2018
An interactive profiler for Python inspired by the Unity3D profiler.

Developed a Python profiler with an interactive TUI inspired by the Unity profiler. On GitHub, this project has been starred by 3k people. Also, it was the 3rd daily trending repository on Sep 22, 2014.

Others
- Tossi – A utility for Korean allomorphic particles.
- Flask-AutoIndex – mod_autoindex for Flask.
- SUBLEERUNKER – A simple parody game of SUBERUNKER.
- Me2virus – An XSS attack on social media named Me2day. When a user looks at an infected post, a new infected post was written on the user's wall.

Contributions
- For PyTorch, fixed potential GPU memory violation (#27371); deprecated inconsistent API (#21006, #25985); discussed a counterintuitive behavior of the autograd engine (#18568).
- For ZeroMQ, discussed a PUB socket crash (#2942).
- For Flask, fixed a bug to generate URL with a subdomain (#108).
- For jQuery 1.4.3, restored a missing part of the content negotiation header for Ajax.

Publications

- C. Kim*, H. Lee*, M. Jeong, W. Baek, B. Yoon, I. Kim, S. Lim, S. Kim. "torchgpipe: On-the-fly Pipeline Parallelism for Training Giant Models". arXiv preprint arXiv:2004.09910. (2020)

    *Contributed equally.

Public Speeches

- Remake of Hangulize[ko] at Golang Korea 2018 and Naver D2
- Server architecture of Durango Vol. 3[ko] at NDC 2018
- Python Survival Guide[ko] at Nexon Talk 2016
- Server architecture of Durango Vol. 2[ko] at NDC 2016
- Profiling[ko] at PyCon KR 2015
- Server architecture of Durango[ko] at NDC 2014

Languages

- Korean – native
- English – conversational

Education

Computer Software, Kwangwoon University, 2008 – Completed the first year only.