Topik : 1.1. Pengenalan Federated Learning

Objective : Memahami konsep dasar FL dan manfaatnya dibanding centralized learning

Task : Tulis ringkasan perbandingan centralized vs federated; jelaskan 2 use case FL

Source :

https://dev.to/alex-nguyen-duy-anh/communication-efficient-learning-of-deep-networks-from-decentralized-data-by-alex-nguyen-3df4

Introduction to Communication-Efficient Learning of Deep Networks from Decentralized Data

Communication-Efficient Learning of deep network merevolusi cara kita melatih model pembelajaran mesin, terutama dalam scenario Dimana privasi data menjadi pertimbangan kritis. Pendekatan ini, yang sering diwujudkan melalui seperti Federated Learning (FL), yang memungkinkan pelatihan jaringan yang canggih tanpa perlu memusatkan data mentah, sehingga secara signifikan mengurangi beban komunikasi sekaligus menjaga privasi pengguna.

Dengan menyimpan data secara local di perangkat seperti ponsel atau system IOT dan hanya berbagi pembaruan model, pelatihan dapat dilakukan secara efisien di seluruh jaringan perangkat terdistribusi. Metode ini secara fundamental mengubah pendekatan pelatihan model dengan memungkinkan penggabungan pengetahuan dari berbagai sumber data tanpa perlu memustkan data itu sendiri.

Pendekatan semacam iini tidak hanya meningkatkan privasi dan mengurangi beban komunikasi, tetapi juga menciptakan lingkungan pembelajaran yang lebih scalable dan inklusif.

Overview of Communication-Efficient Learning

Communication-Efficient Learning berfokus pada pengurangan jumlah data yang ditransfer selama proses pelatihan jaringan saraf dalam (deep neural networks).

Dalam pembelajaran terpusat (centralized learning) , semua data harus di kumpulkan di satu Lokasi, yang seringkali tidak praktis dan tidak efisien karena masalah privasi dan keterbatasan bandwidth.

Sebaliknya, Teknik – Teknik communication-efficient seperti Federated Learning memungkinkan setiap perangkat memproses data lokalnya dan hanya berbagi pembaruan model, sehingga secara drastic mengurangi data yang perlu dikirimkan melalui jaringan.

Metode ini juga mendukung pelatihan model berskala besar di berbagai perangkat yang tersebar secara geografis, memanfaatkan kekuatan komputasi gabungan mereka tanpa mengorbankan privasi pengguna.

Dengan meminimalkan transfer data, Teknik ini tidak hanya menghemat sumber daya jaringan tetapi juga membuat pelatihan menjadi lebih layak di lingkungan dengan konektivitas terbatas atau latensi tinggi.

Definition and Importance of Decentralized Deep Network Training

Pelatihan jaringan dalam (deep network) terdesentralisasi mengacu pada proses Dimana beberapa perangkat atau node berpartisipasi dalam melatih model kolektif, tetapi masing – masing menyimpan datanya secara privat dan hanya berbagi hasil komputasi lokalnya.

Pendekatan ini sangat penting dalam scenario Dimana data tidak dapat dipindahkan atau dibagikan secara hukum maupun etik, seperti pada aplikasi Kesehatan Dimana kerahasiaan pasien harus dijaga, atau di lingkungan seluler Dimana data pribadi pengguna tidak boleh meninggalkan perangkat mereka.

Pentingnya metode ini terletak pada kemampuannya untuk memanfaatkan kekuatan big data sekaligus melindungi privasi individu. Hal ini memungkinkan organisasi untuk melatih model yang kuat pada Kumpulan data yang sebenarnya tidak dapat diakses karena kendala regulasi atau tantangan logistic.

Reduction of Communication Overhead Compared to Centralized Training

Berbeda dengan Centralized Training yang mengharuskan semua data dikumpulkan di satu Lokasi, Communication-Efficient secara signifikan mengurangi jumlah data yang ditransmisikan.

Pengurangan beban komunikasi ini dicapai dengan hanya mengirimkan pembaruan model (model updates) alih-alih keseluruhan Kumpulan data, yang dapat menghasilkan penghematan besar dalam bandwidth dan konsumsi energi.

Sebagai contoh, dalam skenario di mana jutaan perangkat seluler berkontribusi pada pelatihan model, efek kumulatif dari pengurangan komunikasi dapat menghasilkan peningkatan performa dan penghematan biaya yang signifikan.

Emergence of Federated Learning (FL)

Sebagai tekni perintis, FL tidak hanya menjawab kebutuhan mendesak untuk mengurangi beban komunikasi, tetapi juga membuka jalan bagi inovasi lebih lanjut dalam metodologi pelatihan terdesentralisasi.

Kemunculan FL di dorong oleh kebutuhan untuk memanfaatkan data dalam jumlah besar yang dihasilkan di perangkat edge tanpa mengorbankan privasi pengguna.

The Dramatic Communication Reduction

Salah satu aspek paling menarik dari FL Adalah kemampuannya melatih model dengan putaran komunikasi yang jauh lebih sedikit di bandingkan metode tradisional.

Pengurangan dramatis ini sering disebut mencapai 10 – 100 kali lebih sedikit. Secara langsung berdampak pada biaya operasional yang lebih rendah dan waktu penyebaran model yang lebih cepat.

Pengurangan frekuensi komunikasi ini tidak mengorbankan kualitas model. Sebaliknya, hal ini justru menegaskan potensi FL untuk merevolusi cara pengembanngan dan penyebaran model AI dalam scenario praktis, terutama yang membutuhkan pengambilan Keputusan secara real-time.

Addressing Unbalanced and Non-IID Data Distributions at Scale

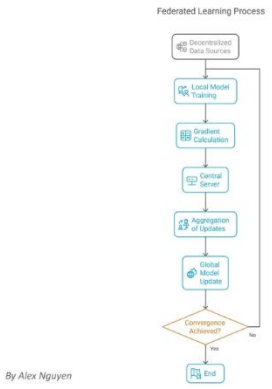
Data di dunia nyata sering kali menunjukkan ketidakseimbangan dan karakteristik non-IID (tidak independen dan tidak terdistribusi secara identik), yang menjadi tantangan besar bagi algoritma pembelajaran tradisional.

FL, dengan sifat terdesentralisasinya, menawarkan solusi menjanjikan karena secara alami dapat mengakomodasi distribusi data seperti ini.

Dengan melatih model pada berbagai kumpulan data dari banyak perangkat kemudian menggabungkan pembaruan ini, FL dapat meningkatkan kemampuan generalisasi dan kinerja model pada berbagai jenis masukan.

Kemampuan ini sangat penting untuk aplikasi yang membutuhkan kinerja kuat di berbagai konteks dan demografi pengguna, menjadikan FL sebagai enabler kunci dalam membangun sistem AI yang benar-benar inklusif.

Fundamentals of Federated Learning (FL)



Pada dasarnya, Federated Learning (FL) beroperasi dengan prinsip melatih model global melalui agregasi pembaruan yang dihitung secara lokal dari banyak klien.

Pendekatan terdesentralisasi ini memastikan data sensitif tetap berada di perangkat yang menghasilkannya, sehingga meningkatkan privasi.

Dengan hanya mengirimkan pembaruan model alih-alih data mentah, FL mencapai efisiensi penggunaan bandwidth yang sangat menguntungkan untuk penerapan skala besar pada perangkat edge yang heterogen.

Landasan pemikiran FL memiliki dua aspek utama:

1. Melindungi privasi pengguna dengan menghindari sentralisasi data
2. Mengoptimalkan pemanfaatan sumber daya komputasi yang tersebar di banyak perangkat

Metode ini memanfaatkan keberadaan perangkat edge yang tersebar luas, mengubahnya dari sekadar penghasil data menjadi partisipan aktif dalam proses pembelajaran.

Perbandingan Centralized Learning vs Federated Learning

Aspek	Centralized Learning	Federated Learning
Data Storage	Semua data dari berbagai sumber di kumpulkan di satu server pusat	Data tetap tersimpan di perangkat lokal (tidak dikirim ke pusat), hanya update model yang dikirim
Privasi	Rentan, karena data mentah yang harus dipindahkan dan disimpan terpusat	Lebih aman, privasi pengguna terjaga karena data tidak keluar dari perangkat.
Bandwidth	Tinggi, karena data mentah dalam jumlah besar harus di transfer / dikirim	Rendah, hanya parameter/model update yang di transfer.
Skalabilitas	Sulit jika data sangat besar dan tersebar di berbagai Lokasi (Terbatas oleh kapasitas server)	Sangat scalable karena memanfaatkan kekuatan komputasi terdistribusi di perangkat.

Source :

https://research.aimultiple.com/federated-learning/

Federated Learning (FL) adalah pendekatan *machine learning* terdesentralisasi yang memungkinkan beberapa organisasi atau perangkat melatih model *machine learning* secara kolaboratif **tanpa berbagi data pribadi**. Alih-alih mengirim data mentah ke server pusat, hanya **pembaruan model** atau parameter model yang dipertukarkan, sehingga menjaga privasi dan keamanan data.

Dengan mempertahankan data pelatihan di perangkat lokal dan hanya menggabungkan *insight*-nya, FL meningkatkan privasi data sekaligus memanfaatkan data terdistribusi untuk meningkatkan akurasi model.

Manfaat :

1. Mendukung *continual learning* melalui data lokal di perangkat pengguna.
2. Data tetap di perangkat, tidak dikirim ke server pusat.
3. Mematuhi regulasi privasi dan mengurangi risiko kebocoran data.

Source : https://mlcommons.org/2023/07/announcing-medperf-open-benchmarking-platform-for-medical-ai/

Kecerdasan Buatan Medis (Medical AI) memiliki potensi luar biasa untuk memajukan layanan kesehatan dan meningkatkan kualitas hidup semua orang di seluruh dunia. Namun, keberhasilan penerapan di klinis membutuhkan evaluasi kinerja model AI pada dataset dunia nyata yang besar dan beragam.

Federated evaluation pada MedPerf: Model pembelajaran mesin didistribusikan ke pemilik data untuk dievaluasi secara lokal di tempat mereka, tanpa perlu mengekstrak data ke lokasi pusat.

MedPerf adalah platform *open benchmarking* yang secara efisien mengevaluasi model AI pada data medis dunia nyata yang beragam dan memberikan efektivitas klinis, sambil tetap memprioritaskan privasi pasien serta mengurangi risiko hukum dan regulasi. **MedPerf** meningkatkan AI Medis dengan membuat data medis di seluruh dunia dapat diakses oleh peneliti AI secara aman dan efisien, sehingga mengurangi bias dan meningkatkan generalisasi serta dampak klinis.

Secara kritis, **MedPerf** memungkinkan organisasi kesehatan untuk menilai dan memvalidasi model AI melalui proses efisien dengan pengawasan manusia tanpa harus mengakses data pasien. Desain platform ini mengandalkan **federated evaluation**, di mana model AI medis dikirimkan secara jarak jauh dan dievaluasi di lokasi penyedia data. Pendekatan ini mengatasi kekhawatiran terkait privasi data dan membangun kepercayaan di antara pemangku kepentingan layanan kesehatan, sehingga menciptakan kolaborasi yang lebih efektif.

Use Case 1 : MedPerf – Federated Benchmarking Platform for Medical AI

Latar Belakang :

Model AI medis biasanya dilatih dengan data dari lingkungan klinis yang terbatas dan spesifik, yang dapat menyebabkan bias yang tidak disengaja terhadap populasi pasien tertentu. Kurangnya generalisasi ini dapat mengurangi dampak nyata AI Medis. Namun, mengakses dataset yang lebih besar dan beragam sulit dilakukan karena pemilik data dibatasi oleh risiko privasi, hukum, dan regulasi.

Solusi :

MedPerf adalah platform *open benchmarking* yang menggunakan Federated Evaluation untuk mengevaluasi model AI medis:

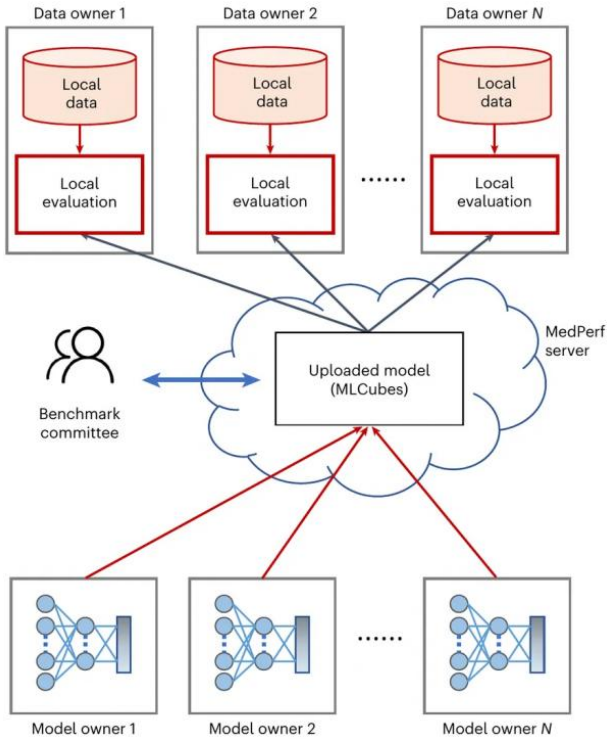
- Data pasien tetap berada di rumah sakit masing-masing (tidak dipindahkan).
- Model AI dikirimkan ke rumah sakit untuk dijalankan secara lokal pada data mereka.
- Hanya hasil evaluasi/pembaruan model yang dibagikan kembali.
- Memastikan privasi pasien tetap terlindungi sambil meningkatkan keakuratan model secara global.

Studi Kasus Implementasi

- **Federated Tumor Segmentation (FeTS) Challenge**
    - Melibatkan **32 rumah sakit di 6 benua**.
    - 41 model AI dievaluasi menggunakan MedPerf.
    - Hasil: validasi AI tumor otak yang lebih inklusif, cepat, dan tetap sesuai regulasi privasi.

Manfaat Utama

1. **Privasi Terjaga** → Data medis sensitif tidak pernah keluar dari institusi kesehatan.
2. **Mengurangi Bias** → Model divalidasi pada populasi pasien yang beragam secara global.
3. **Efisiensi Waktu & Biaya** → Evaluasi model dapat dilakukan dalam hitungan jam, bukan berbulan-bulan.
4. **Kolaborasi Global** → Mendukung penelitian multi-negara, multi-institusi tanpa hambatan regulasi data.



Evaluasi terdesentralisasi pada MedPerf. Model pembelajaran mesin didistribusikan kepada pemilik data untuk dievaluasi secara lokal di lokasi mereka tanpa perlu atau kewajiban untuk mengekstrak data mereka ke lokasi pusat.

Source : https://arxiv.org/pdf/2202.01141

Use Case 2 : FLDDPG – Federated Reinforcement Learning untuk Navigasi Kolektif Swarm Robotik

Latar Belakang :

Kemajuan terbaru dalam Deep Reinforcement Learning (DRL) telah meningkatkan bidang robotika dengan memfasilitasi desain pengendali otomatis, yang sangat penting untuk sistem robotika swarm. Sistem ini memerlukan pengendali yang lebih canggih daripada konfigurasi robot tunggal untuk mencapai perilaku kolektif yang terkoordinasi.

Meskipun desain pengendali berbasis DRL terbukti efektif, ketergantungannya pada server pelatihan pusat menimbulkan tantangan di lingkungan dunia nyata dengan komunikasi yang tidak stabil atau terbatas.

Solusi :
FLDDPG menggabungkan Federated Learning (FL) dengan algoritma DRL Deep Deterministic Policy Gradient (DDPG):

- Robot swarm melakukan pelatihan model secara lokal dengan data sensor dan interaksi mereka sendiri.
- Model yang dilatih dikirim ke server pusat untuk agregasi model global.
- Model global kembali disebarkan ke robot sebagai pembaruan. Pendekatan ini mengurangi kebutuhan bandwidth tinggi, menjaga privasi data sensor lokal, dan memungkinkan adaptasi terhadap karakteristik spesifik tiap robot.