Evento	Salão UFRGS 2022: SIC - XXXIV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2022
Local	Campus Centro - UFRGS
Título	Lidando com stragglers usando planos de dados programáveis
Autor	DIEGO CARDOSO NUNES
Orientador	ALBERTO EGON SCHAEFFER FILHO

A possibilidade de usar diversos computadores para distribuir tarefas de *Machine Learning* (ML) vem sendo explorada para permitir o treinamento de modelos cada vez mais complexos. Uma estratégia comum é o paralelismo de dados, na qual múltiplos trabalhadores treinam localmente instâncias de um mesmo modelo sobre conjuntos de dados diferentes. Esses modelos locais são então agregados para criar um modelo atualizado, que é enviado de volta aos trabalhadores. Pesquisas recentes propuseram usar a rede para acelerar a agregação usando planos de dados programáveis. Entretanto, esses trabalhos apenas aplicaram estratégias de agregação síncronas, que sofrem severamente com a presença de *stragglers* (trabalhadores lentos), e assíncronas, que podem sofrer com problemas de convergência. Surge então a possibilidade de explorar estratégias híbridas como forma de lidar com *stragglers,* buscando terminar o treinamento mais rapidamente e causar um impacto menor na convergência do modelo. Sendo assim, o objetivo do projeto é desenvolver um sistema de agregação em rede capaz de lidar com *stragglers*, inspirado por estratégias de sincronização híbridas, como a *Stale Synchronous Parallel*. Para tal, realizou-se uma pesquisa bibliográfica sobre o estado da arte da agregação em rede para treinamento de ML distribuído. Com base nisso, propôs-se uma nova estratégia de agregação em rede, baseada no gerenciamento de barreiras de sincronização no plano de dados através do controle de relógios lógicos. A pesquisa ainda está em andamento, tendo como resultados parciais o protótipo do sistema, desenvolvido para o *software switch* BMv2 usando a linguagem P4. Para esse protótipo, foram desenvolvidos e implementados algoritmos e estruturas de dados com uso de memória eficiente, tendo em vista as limitações do modelo de programação do plano de dados. O sistema será avaliado usando cargas reais de treinamento de ML, com foco no treinamento de redes neurais.