

Optimization des Hyperparamètres appliquée au Fine Tuning de LLM

Basé sur l'article : *Bayesian and Partition-Based Optimization for Hyperparameter Optimization of LLM Fine-Tuning*

Nathan Davouse

Optimisation des Hyperparamètres (OHP)

Hyperparamètres

Paramètres qui ne sont pas entraînés par le modèle
(learning rate, dropout ...)

Objectifs

- ▶ Meilleure performance qu'en manuel
- ▶ Retirer le besoin d'expertise

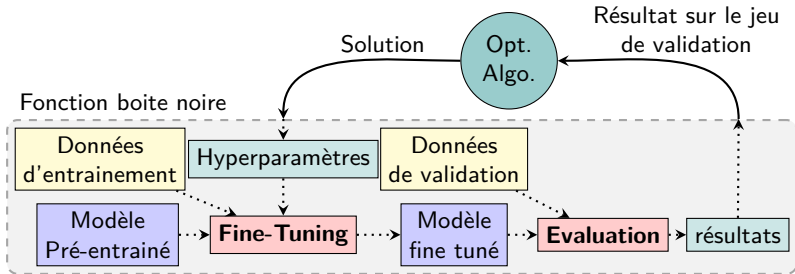


Figure: Fonctionnement général de l'optimisation des hyperparamètres

Travaux connexes

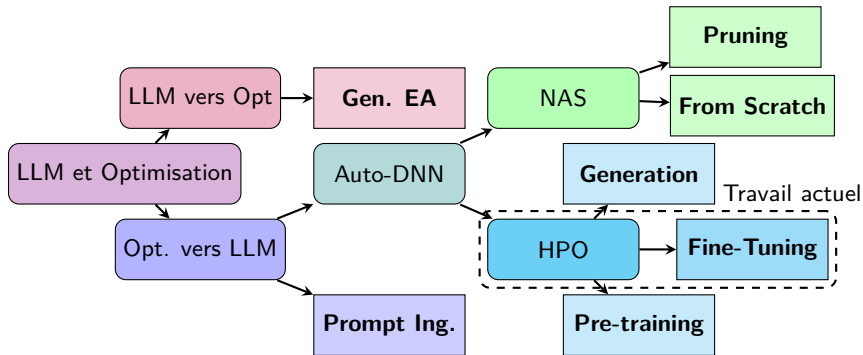


Figure: Classification des travaux similaires

Sommaire

1. Introduction

2. Design et Implémentation

3. Résultats et Analyses

4. Conclusion

Stratégie de Recherche : Bayesian Multi-Scale Optimistic Optimization (BaMSOO)

Décomposition suivant SOO, mais utilisant des Process Gaussien pour éviter les évaluations non prometteuses.

Evaluation par BaMSOO

- ▶ If $UCB(x) > f^+$:
 - $g(x) = f(x)$ real evaluation
- ▶ Else :
 - $g(x) = LCB(x)$ use LCB to replace $f(x)$

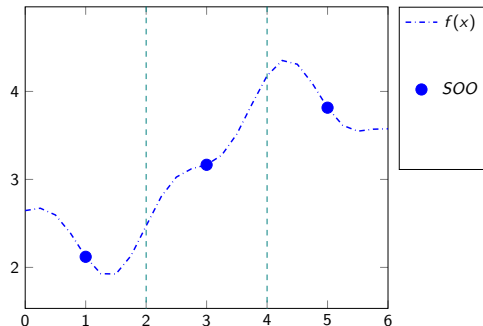


Figure: Illustration de l'Algorithme BaMSOO

Stratégie de Recherche : Bayesian Multi-Scale Optimistic Optimization (BaMSOO)

Décomposition suivant SOO, mais utilisant des Process Gaussien pour éviter les évaluations non prometteuses.

Evaluation par BaMSOO

- ▶ If $UCB(x) > f^+$:
 - $g(x) = f(x)$ real evaluation
- ▶ Else :
 - $g(x) = LCB(x)$ use LCB to replace $f(x)$

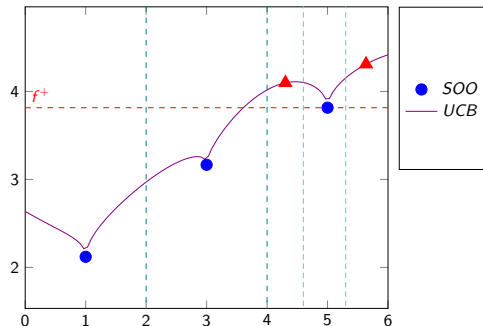


Figure: Illustration de l'Algorithme BaMSOO

Sommaire

3. Résultats et Analyses

Résultats des 3 algorithms

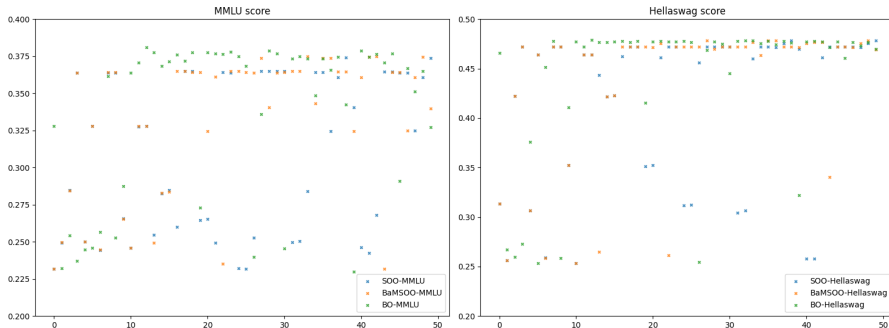


Figure: Résumé des résultats par sampling. Détails par algorithme dans les annexes 5, 6 et 7

Perspectives

Poursuite du travail

- ▶ Retour sur l'article et présentation en conférence (si validation)
- ▶ Elargissement de l'espace de recherche
- ▶ Diversification sur les modèles/données

Généralisation hors LLM

- ▶ Parallélisation d'algorithme d'optimisation de fonction coûteuse pour Exascale (project Exa-MA)
- ▶ Intégration de substitut dans les algorithme parallèle à Partition ³

³Partition-based Parallel Bayesian Optimisation

Sommaire

1. Introduction

2. Design et Implémentation

3. Résultats et Analyses

4. Conclusion

Résultats du stage

Une conclusion

Apprentissage

Une conclusion

Poursuite du projet professionnel

Une conclusion

Merci.

Annexe 1 : Stratégie d'Evaluation de Solutions

Implémentation

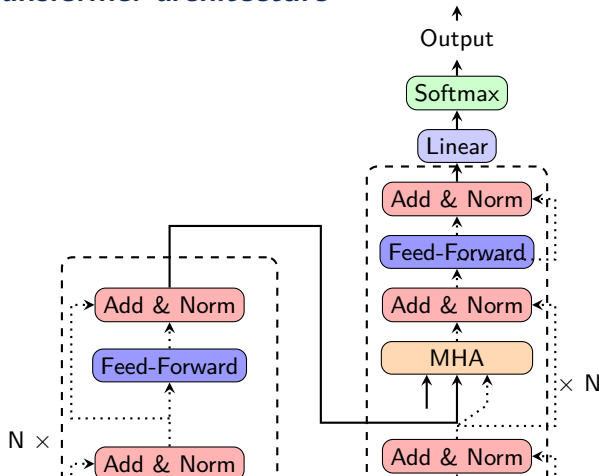
► Fine Tuning

- Modèle : LLaMa-3.2-1B
- Jeu de données d'entraînement : Alpaca
- LitGPT framework : basé sur Pytorch, facilite le Fine Tuning de LLM

► Evaluation

- librairie lm_eval : standard pour l'évaluation de LLM
- Evaluation par la précision sur des jeu de données Benchmark : Hellaswag et MMLU

Annexe 2 : Transformer architecture



Annexe 3 : Multi-Head Attention

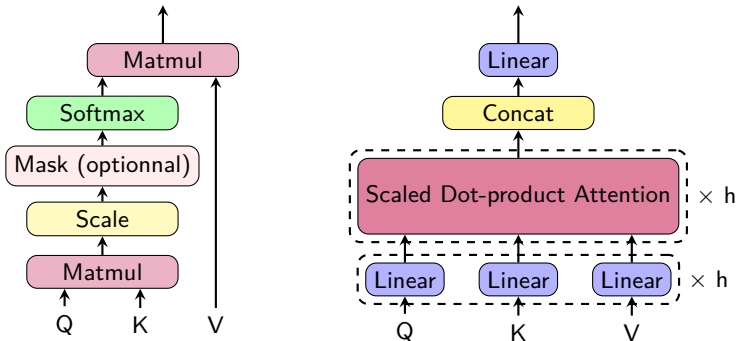


Figure: Illustration du mécanisme d'auto-attention : A droite le mécanisme complet, à gauche le *Scaled Dot-product Attention*

Annexe 4 : Low Rank Adaptation (LoRA)

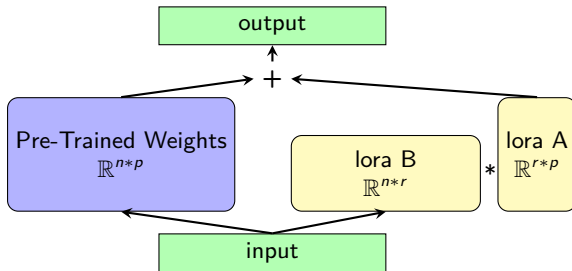


Figure: Illustration de l'application du Low Rank Adaptation (LoRA)

Annexe 5 : Résultats pour BO

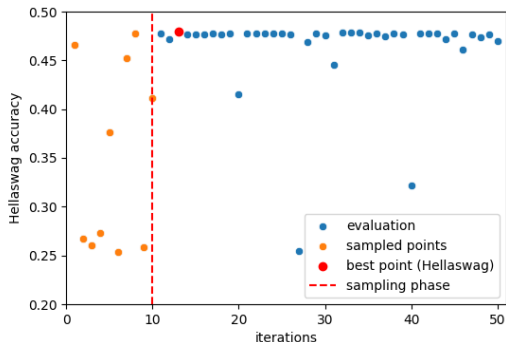


Figure: Evolution des score lors de l'expérience BO

Annexe 6 : Résultats pour S00

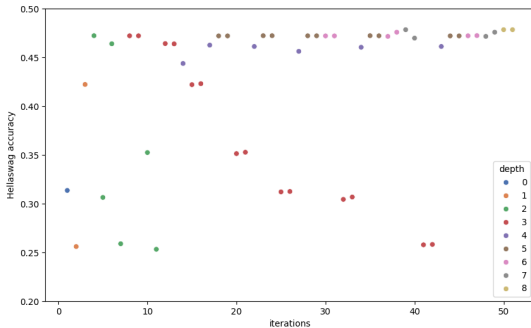


Figure: Evolution des score lors de l'expérience SOO

Annexe 7 : Résultats pour BaMSOO

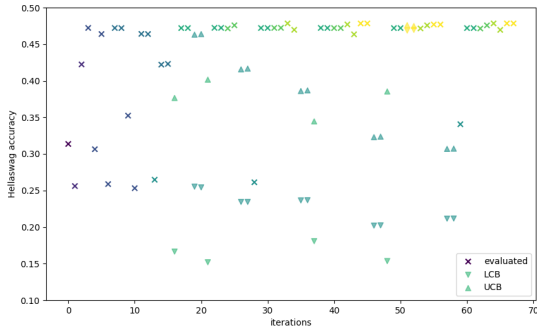


Figure: Evolution des score lors de l'expérience BaMSOO

Annexe 8 : Modèle Beamer UTT



beamer_utt_template

Public



A template for presentation in latex with the graphical chart of University of Technology of Troyes

● TeX

Figure: Presentation du template beamer sur github : lien