COMPTE RENDU TD°3

Intelligence Artificielle & Optimisation

DUNAND Tom - 5A ILIA

Travail Réalisé

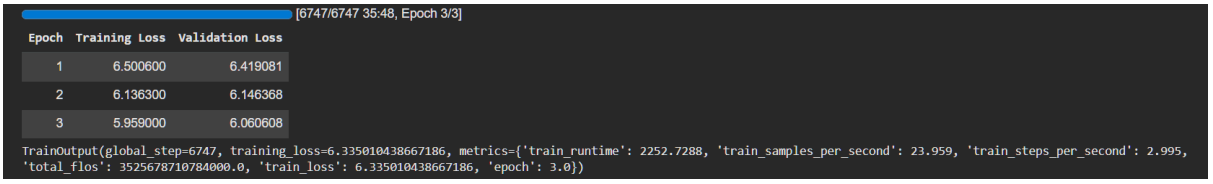
Le dataset WikiText-2 contient plus de 2 millions de tokens d'articles Wikipédia. Composé de 36 000 données d'entraînement, 3 700 de validation et 4 300 de test.

Étapes du notebook :

- Installation de transformers et datasets, chargement de WikiText-2
- Entraînement d'un tokenizer (BPE qui découpe le textes et les mots intelligemment) qui découpe en 52 000 tokens sur le dataset
- Regroupement des textes en blocs de 128 tokens pour l'entraînement
- 2 modèles instanciés, même architecture de base mais différents types d'attention et deux objectifs différents :
    - Modèle Causal LM (GPT-2) : Pour la génération de texte (12 couches, 768 dimensions, 12 têtes d'attention)
    - Modèle Masked LM (RoBERTa) : Pour le masquage de tokens (architecture similaire)
- 5. Entraînement : 3 époques avec l'API Trainer (learning rate 5e-5 —> avec scheduler, batch size=4)
- 6. Évaluation et génération : Test de perplexité (plus la perplexité est élevée plus les prédictions du modèle sont confuse donc moins le modèle est bon) et génération de texte
- 3 époques avec learning rate 5e-5, batch size 4, warmup de 500 steps
- Test de génération de texte avec le modèle entraîné

Résultats :

CLM (gpt-2) :



MLM (BERT) :

Epoch	Training Loss	Validation Loss
1	7.091300	7.054184
2	6.899600	6.889086
3	6.848700	6.870905

TrainOutput(global_step=7038, training_loss=7.047292684959938, metrics={'train_runtime': 2112.6151, 'train_samples_per_second': 26.641, 'train_steps_per_second': 3.331, 'total_flos': 3703423157830656.0, 'train_loss': 7.047292684959938, 'epoch': 3.0})

Observations

- La loss diminue régulièrement sur les 3 époques donc l'apprentissage est efficace
- Pas de surapprentissage observé (validation loss proche de training loss)
- Temps d'entraînement : 30 minutes sur GPU T4 gratuit de google Colab
- GPT-2 est plus performant que BERT sur cette tâche

Commentaires

L'utilisation du framework HuggingFace était intéressante à approfondir. Il est aussi intéressant de voir l'impact important qu'a le type d'attention du modèle sur les prédictions et la perplexité.