Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

DivMerge: une méthode de fusion de modèles pour le multi-tâches fondée sur une divergence

Brahim Touayouch[1, 2] Loïc Fosse[1, 3] Géraldine Damnati[1] Gwénolé Lecorvé[1]
(1) Orange Research, 2, avenue Pierre Marzin, 22300 Lannion, France
(2) École polytechnique, Institut polytechnique de Paris, Route de Saclay, 91120 Palaiseau, France
(3) CNRS, LIS, Aix Marseille Université, 163 avenue de Luminy, 13288 Marseille, France
prenom.nom@orange.com

RÉSUMÉ

La fusion de modèles affinés est une alternative prometteuse à un entraînement multi-tâches classique par mélange de données. Cependant, les possibles interférences entre tâches constituent un frein, surtout à mesure que le nombre de tâches à fusionner augmente. Nous présentons DivMerge, une méthode qui fusionne des modèles affinés sur différentes tâches en minimisant la divergence de Jensen-Shannon entre leurs sorties et celles du modèle fusionné, ceci sans données annotées et en équilibrant automatiquement l'importance respective de chaque tâche. Outre de solides propriétés théoriques démontrées par notre méthode, nos expériences sur des tâches de classification et de génération avec des modèles auto-régressifs montrent que DivMerge surpasse systématiquement les méthodes de la littérature et est robuste à l'augmentation du nombre de tâches.

ABSTRACT

DivMerge : A divergence-based model merging method for multi-tasking.

Merging fine-tuned models is a promising alternative to costly multi-task training, but task interference remains a challenge, especially as the number of tasks grows. We present DivMerge, a reference-free method that merges models trained on different tasks by minimizing Jensen-Shannon divergence between their outputs and those of the merged model, automatically balancing task importance. While the method exhibits strong theoretical properties, experiments on classification and generative tasks with autoregressive models show that DivMerge consistently outperforms prior work, and remains robust when scaling to more tasks.

MOTS-CLÉS : Fusion de modèles, modèles de langue, divergence.

KEYWORDS: Model merging, language model, divergence.

ARTICLE ACCEPTÉ À : The 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026).
URL : https://arxiv.org/abs/2509.02108

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 13–13.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.