

Normalisation Contextuelle : Une Nouvelle Approche pour la Stabilité et l'Amélioration des Performances des Réseaux de Neurones

Bilal Faye*, Hanane Azzag**, Mustapha Lebbah***, Fangchen Fang****

*LIPN, UMR CNRS 7030, Université Sorbonne Paris Nord, 93430 Villetaneuse, France
faye@lipn.univ-paris13.fr

**LIPN, UMR CNRS 7030, Université Sorbonne Paris Nord, 93430 Villetaneuse, France
azzag@univ-paris13.fr

***Laboratoire DAVID, Université Paris Saclay, 78035 Versailles, France
mustapha.lebbah@uvsq.fr

****Laboratoire L2TI, Université Sorbonne Paris Nord, 93430 Villetaneuse, France
fangchen.feng@univ-paris13.fr

Résumé. L'apprentissage des réseaux de neurones est confronté à des défis majeurs liés au changement de distribution en couches, perturbant ainsi la convergence et les performances des modèles. La Normalisation par lot (BN) a révolutionné ce domaine, mais repose sur l'hypothèse simplifiée d'une seule composante gaussienne par lot. Pour remédier à cela, la Normalisation par Mélange (MN) a adopté une approche basée sur le modèle de mélange gaussien (GMM), mais avec des coûts computationnels importants liés à l'algorithme Espérance-Maximisation (EM) pour déterminer des composantes. Notre solution, la Normalisation Contextuelle (CN), regroupe des observations similaires en "contextes" pour une représentation locale, sans nécessiter d'algorithme de construction de ces contextes. Les paramètres de normalisation sont appris de manière similaire aux poids du modèle, assurant rapidité, convergence et performances supérieures par rapport à BN et MN.

1 Introduction

La normalisation est une opération courante en traitement des données, impliquant des transformations pour obtenir une représentation avec des propriétés statistiques spécifiques Kessy et al. (2018). Elle est utilisée pour égaliser les amplitudes des variables, accélérant ainsi la convergence dans les réseaux de neurones monocouches LeCun et al. (2002). Cependant, dans les réseaux multicouches, la normalisation des entrées peut ne pas avoir d'impact significatif sur les autres couches, car les distributions de données évoluent avec les mises à jour des poids du réseau. Pour garantir la stabilité de l'apprentissage à travers toutes les couches, diverses techniques de normalisation sont nécessaires, notamment la normalisation des activations, des poids et des gradients. La normalisation par lot (batch) (BN) Ioffe et Szegedy (2015) est une méthode courante pour stabiliser l'apprentissage dans les réseaux de neurones