

Prise en Compte du Contexte pour Contraindre les Réseaux Profonds: Application à l'Étiquetage de Scènes

Taygun Kekeç¹, Rémi Emonet¹, Elisa Fromont¹, Alain Trémeau¹, and Christian Wolf²

¹Université de Lyon, CNRS UMR 5516, Laboratoire Hubert-Curien, Université de Saint-Etienne, F-42000, Saint-Etienne, France

²Université de Lyon, CNRS INSA-Lyon, LIRIS, UMR5205, F-69622, France

Résumé

L'apprentissage automatique à l'aide de réseaux profonds est un problème difficile : la complexité du réseau et la méthode de descente de gradient utilisée pour mettre à jour les poids du réseau peuvent tous deux mener à des phénomènes de sur-apprentissage et des mauvais optima locaux pour ces poids. Pour aborder ce problème dans un contexte de classification de scènes (où le but est de classer tous les pixels d'une image), nous cherchons à contraindre certaines couches du réseau en utilisant une partie de l'information sémantique contextuelle disponible pendant la phase d'apprentissage. Nous voulons : 1) contraindre la capacité du réseau (son nombre de paramètres) tout en autorisant l'apprentissage de fonctions complexes et 2) obtenir des couches avec une sémantique plus précise permettant d'éviter les mauvais optima. Nous proposons dans un premier temps d'apprendre un réseau de convolution "faible" qui nous fournirait une première estimation de la classe des étiquettes des pixels qui entourent un pixel donné. Puis, nous incorporons ce réseau faible dans un réseau plus complexe dont les poids ont déjà été préalablement initialisés en prenant en compte les étiquettes des pixels qui entourent notre pixel d'intérêt. Ce processus, qui complexifie itérativement le réseau de convolution, a pour but d'augmenter l'interprétabilité des couches du réseau en contraignant les "feature maps" à encoder une information contextuelle précise. Nous montrons que cette connaissance contextuelle permet d'obtenir des performances légèrement meilleure que l'état de l'art pour les jeux de données les plus connus en étiquetage de scènes que sont "Stanford" et "Sift Flow". Cette approche est, en outre, générique et pourrait être appliquée à tous types de réseaux pour lesquels des informations contextuelles sont disponibles pendant la phase d'apprentissage.

prentissage.

Keywords : Réseaux de neurones à convolution, Classification itérative, Étiquetage de scènes

Remerciements

Ce travail a bénéficié de l'aide de l'ANR dans le cadre du projet SoLStiCe (ANR-13-BS02-0002-01).