Titre : Contribution au développement des méthodes avancées de la programmation DC et DCA pour certaines classes de problèmes d'optimisation non-convexes. Applications en apprentissage automatique.

Résumé :

Ce mémoire est consacré aux méthodes avancées de la programmation DC (**D**ifference of **C**onvex functions) et DCA (DC Algorithm) et leurs applications en apprentissage automatique. La programmation DC et DCA jouent un rôle central dans l'optimisation non-convexe car la plupart des problèmes non-convexes peuvent être formulés/reformulés sous forme d'un programme DC. Après plus de 35 ans de développement, DCA est indéniablement reconnu comme un des rares algorithmes capables de résoudre les problèmes non-convexes non-différentiables de très grande taille. Les travaux sur la programmation DC et DCA présentés dans cette partie s'insèrent dans la nouvelle direction de recherche de DCA, à savoir, développer les nouvelles variantes de DCA pour répondre à plusieurs questions ouvertes importantes: l'amélioration de la vitesses de convergence de DCA, la recherche des bonnes décompositions DC, le développement des schémas DCA adaptés aux problèmes avec big data, l'utilisation de DCA sans mettre en évidence une décomposition DC de la fonction objectif, l'extension de DCA à la résolution des programmes non convexes au-delà de la programmation DC.

Le mémoire s'articule en trois parties. La première partie est dédiée au développement des nouvelles variantes de DCA pour certaines classes de problèmes d'optimisation non-convexes, la deuxième partie se concentre sur l'optimisation parcimonieuse tandis que la troisième partie est consacrée aux méthodes DCA pour le clustering.

La première approche avancée de DCA, nommée Accelerated DCA (ADCA), est pour but d'améliorer la vitesse de convergence de DCA. Nous introduisons la technique d'accélération de Nesterov au schéma DCA standard. Nous démontrons la convergence de ADCA, et estimons le taux de convergence de ADCA sous les conditions de Lojasiewics. Nous déployons ensuite ce schéma ADCA à la résolution une classe spéciale des problèmes souvent rencontrés en pratique, qui est la minimisation de la somme d'une fonction différentiable à dérivé Lipschitzien (possiblement non-convexe) et une fonction DC.

La deuxième approche avancée de DCA, appelée DCA-Like, permet de résoudre deux classes des problèmes sans mettre en évidence une décomposition DC de la fonction objectif : la minimisation de la somme d'une fonction non-convexe différentiable à dérivé Lipschitzien et d'une somme des fonctions composites et la minimisation de la somme d'une fonction différentiable à dérivé Lipschitzien et une fonction DC. DCA-Like est « similaire » à DCA dans le sens où ils approximent itérativement le programme DC original par une séquence des programmes convexes. Cependant, DCA-Like est « différent » de DCA dans la façon de décomposer la fonction objectif qui n'est pas forcément une décomposition DC. Nous montrons que la séquence bornée générée par DCA-Like converge toujours vers un point critique du programme considéré, sous certaines conditions. De plus, nous démontrons que, sous l'hypothèse de Kurdyka-Lojasiewicz, le taux de convergence de DCA-Like est au moins sous-linéaire.

La troisième approche, appelée DCA Stochastique (SDCA en bref), est développée pour faire face aux problèmes avec big data. Nous considérons le problème de minimisation de la somme d'un grand nombre de fonctions DC, e.g. $\min F(x) = \sum_{i=1}^{N} F_i(x)$ où $F_i(x) = h_i(x) - g_i(x)$ sont des fonctions DC et N est un grand nombre, qui apparait dans l'optimisation stochastique ou dans les applications en apprentissage automatique. SDCA, qui nécessite le calcul d'un faible nombre de sous-différentiel de $h_i(x)$ à chaque itération, permet de réduire considérablement le temps de calcul de DCA. Nous démontrons que SDCA converge vers un point critique avec probabilité 1.

La deuxième partie de ce mémoire est consacrée à l'optimisation parcimonieuse et ses applications en apprentissage automatique. L'optimisation parcimonieuse, e.g. la minimisation de la norme zéro, trouve ses applications dans plusieurs domaines comme apprentissage automatique, traitement d'image, finance, etc. Nous présentons deux approches majeures basées sur la programmation DC et DCA pour l'optimisation parcimonieuse. La première approche consiste à reformuler, de manière équivalente, la minimisation de la norme zéro par un problème d'optimisation combinatoire avec les variables binaires. Ce dernier est ensuite reformulé comme un problème DC par la technique de pénalité exacte et résolu par DCA. Dans la deuxième approche, nous remplaçons la norme zéro par les approximations non-convexes. En démontrant que la plupart des approximations non-convexes dans la littérature sont des fonctions DC, nous les unifions et proposons trois algorithmes DCA pour l'optimisation parcimonieuse. Nous fournissons également plusieurs études théoriques sur la consistance de la reformulation et la convergence de nos algorithmes. Nous présentons ensuite trois applications de l'optimisation parcimonieuse en apprentissage automatique : la sélection de variables dans SVM et Semi-Supervisé SVM, l'acquisition comprimée.

La troisième partie de ce mémoire concerne un sujet fondamental de l'apprentissage et fouille de données: le clustering. Nous présentons nos méthodes DCA et DCA-Like pour plusieurs problèmes en clustering : la MSSC (Minimum of Sum of Squares Clustering) avec formulation avec variable binaire de; la formulation à deux niveaux de MSSC avec un noyau Gaussien ; le MSSC avec la pondération de variables ; le clusterings par bloc ; le modèle de mélange gaussien et le clustering des séries temporelles.