

SVM incrémental et parallèle sur GPU

François Poulet*, Thanh-Nghi Do**, Van-Hoa Nguyen***

*IRISA-Textmex
Campus de Beaulieu, 35042 Rennes Cedex
francois.poulet@irisa.fr
http://www.irisa.fr/textmex/people/poulet/index_fr.php

** Dpt LUSSE, Télécom Bretagne
Technopôle Brest-Iroise CS 83818, 29238 Brest Cedex 3
tn.do@telecom-bretagne.eu
<http://perso.enst-bretagne.fr/tndo>

***IRISA Symbiose
Campus de Beaulieu, 35042 Rennes Cedex
vhnguyen@irisa.fr
<http://www.irisa.fr/symbiose/old/people/nguyen/>

Résumé. Nous présentons un nouvel algorithme incrémental et parallèle de Séparateur à Vaste Marge (SVM ou Support Vector Machine) pour la classification de très grands ensembles de données en utilisant le processeur de la carte graphique (GPUs, Graphics Processing Units). Les SVMs et les méthodes de noyaux permettent de construire des modèles avec une bonne précision mais ils nécessitent habituellement la résolution d'un programme quadratique ce qui requiert une grande quantité de mémoire et un long temps d'exécution pour les ensembles de données de taille importante. Nous présentons une extension de l'algorithme de Least Squares SVM (LS-SVM) proposé par Suykens et Vandewalle pour obtenir un algorithme incrémental et parallèle. Le nouvel algorithme est exécuté sur le processeur graphique pour obtenir une bonne performance à faible coût. Les résultats numériques sur les ensembles de données de l'UCI et Delve montrent que notre algorithme incrémental et parallèle est environ 70 fois plus rapide sur GPU que sur CPU et significativement plus rapide (plus de 1000 fois) que les algorithmes standards tels que LibSVM, SVM-perf et CB-SVM.

1 Introduction

Les algorithmes de Séparateurs à Vaste Marge proposés par (Vapnik, 1995) et les méthodes de noyaux permettent de construire des modèles précis et deviennent des outils de classification de données de plus en plus populaires. On peut trouver de nombreuses applications des SVM comme la reconnaissance de visages, la catégorisation de textes ou la bioinformatique (Guyon, 1999). Cependant, les SVM demandent la résolution d'un