

Plans d'expériences optimaux pour approximer des surfaces de réponse avec des SVMs

Benoît Gandar

Cemagref de Clermont-Ferrand

LISC (Laboratoire d'Ingénierie pour les Systèmes Complexes)

24, avenue des Landais - BP 50 085 - 63 172 Aubière Cedex 1

Tél : 04 73 44 06 95 - E-Mail : benoit.gandar@cemagref.fr

<http://www.lisc.clermont.cemagref.fr>

Mon travail, qui se situe dans le cadre du projet européen PATRES (<http://www.patres-project.eu>), vise à approximer des noyaux de viabilité de systèmes où ces noyaux sont considérés comme des surfaces de réponse de modèles. Ces surfaces de réponse sont formées par l'ensemble des points configurant certaines propriétés au système. Mon laboratoire a développé (voir [2] et [4]) une méthode d'approximation de ces noyaux par classification¹ en utilisant l'algorithme des SVMs (Séparateur à Vaste Marge ou Machine à Vecteur de Support, [3] et [11]), celui-ci étant particulièrement adapté au problème. Dans ce cadre, notre approche consiste d'abord à générer les plans d'expériences optimaux pour l'apprentissage statistique. Par la suite, nous pourrions nous intéresser à la détermination des points les plus informatifs comme le fait Feuillard (voir [5]) dans le cas de la calibration d'un code de calcul sur une base de données.

En s'appuyant sur les travaux sur les approximations d'intégrales, Cervellera & Muselli montrent dans [1] que les bornes théoriques d'erreur en généralisation sont meilleures en utilisant des bases d'apprentissage générées par des suites à discrétance² faible plutôt qu'en utilisant des tirages aléatoires uniformes dans le cadre de l'approximation de fonctions. Cette borne est obtenue en appliquant le théorème de Koksma-Hlawka. En notant f la fonction à apprendre, \hat{f} la fonction apprise avec un algorithme d'apprentissage sur un ensemble X , $D^*(X)$ la discrétance de celui-ci, le théorème établit :

$$\text{Erreur d'apprentissage} = \int |f - \hat{f}| \leq V_{HK}(|f - \hat{f}|) D^*(X)$$

où $V_{HK}(g)$ représente la variation au sens de Hardy-Krause de la fonction g : une mesure de la régularité de celle-ci. Dans cette inégalité, il apparaît clairement que l'erreur réalisée est directement proportionnelle à la discrétance de l'ensemble d'apprentissage (en supposant que la variation soit finie). L'hypothèse de variation finie, qui est une hypothèse forte, est étudiée et affaiblie par Mary dans sa thèse dans le cas de la régression (voir [9]).

La démarche théorique permettant d'obtenir les bornes d'erreur en généralisation pour l'approximation de fonctions n'est pas adaptée à l'apprentissage de la surface de réponse (apprentissage de la frontière d'une variété). C'est un peu surprenant, car la classification peut apparaître comme un cas particulier du problème général d'approximation de fonctions. En effet, la borne supérieure de la précédente inégalité vaut l'infini, et ceci en raison de la variation infinie des fonctions indicatrices (voir [10]).

¹En Français, le terme correct serait plutôt discrimination que classification qui est le terme Anglais.

²La discrétance d'une suite définie par Niederreiter dans [8] peut être vue comme une mesure d'une répartition des points uniforme dans tous les sous-ensembles de l'ensemble considéré.

A défaut de disposer d'une approche théorique satisfaisante, l'intérêt des suites à discrédance faible dans l'apprentissage en classification peut être exploré expérimentalement. Ainsi, Iwata & Ishii [7] ont observé expérimentalement un gain de qualité dans la classification avec des réseaux de neurones.

Des tests numériques, réalisés avec l'algorithme des SVMs, ne permettent pas de confirmer ces observations. Comparant les apprentissages réalisés sur des échantillons de taille identique issus soit de grilles régulières, soit de suites à discrédance faible, nous avons remarqué dans [6] que les SVMs obtenus à partir d'échantillons issus de suites possèdent des qualités de généralisation légèrement inférieures aux autres.

Ceci nous amène à supposer que le critère fondé sur la discrédance utilisé pour estimer la qualité de l'ensemble d'apprentissage n'est pas le critère optimum. Nous proposons des premiers résultats théoriques, complétés par des expériences numériques sur la dispersion, c'est-à-dire la caractérisation du plus grand vide de l'ensemble, comme caractère fondamental de l'ensemble d'apprentissage pour caractériser la qualité d'apprentissage en classification.

Références

- [1] CERVELLERA, C. and MUSELLI, M. (2004) : Deterministic Design for Neural Network Learning : An Approach Based on Discrepancy. IEEE Transactions on neural network, vol. 15, No 13 , pp. 533-544, May 2004.
- [2] CHAPEL, L. and DEFFUANT, G. (2007) : SVM viability controller active learning : application to bike control. In IEEE Approximate Dynamic Programming and Reinforcement Learning. Hawaï, États-Unis, april 1-5, 2007.
- [3] CORNUEJOLS, A. (2002) : Une nouvelle méthode d'apprentissage : Les SVM. Séparateurs à vaste marge. In Bulletin de l'AFIA, No 51, juin 2002.
- [4] DEFFUANT, G. and CHAPEL, L. and MARTIN, S. (2007) : Approximating viability kernels with Support Vector Machines. In IEEE transactions on automatic control, 52(5), 2007, 933-937.
- [5] FEUILLARD, V. (2007) : Analyse d'une base de données pour la calibration d'un code de calcul, Thèse de doctorat de l'Université Pierre et Marie Curie, et CEA (Commissariat à l'Énergie Atomique).
- [6] GANDAR, B. and DEFFUANT, G. and LOOSLI, G. (2008) : Les suites à discrédance faible : un moyen de réduire le nombre de vecteurs supports des SVMs ? 12^{ième} Journée Scientifique de l'Ecole Doctorale SPI : Apprentissage statistique - Apprentissage symbolique. Annales scientifiques de l'Université Blaise Pascal, Clermont-Ferrand II.
- [7] IWATA, K. and ISHII, N. (2002) : Discrepancy as a Quality Measure for Avoiding Classification Bias. In Proceedings of the 2002 IEEE International Symposium on Intelligent Control. Vancouver, Canada, October 27-30, 2002.
- [8] NIEDERREITER, H. (1992) : Random Number Generation and Quasi-Monte Carlo Methods, Ed. Society for Industrial and Applied Mathematics.
- [9] MARY, J. (2005) : Étude de l'Apprentissage Actif, Application à la Conduite d'Expériences, Thèse de Doctorat de l'Université Paris XI.
- [10] OWEN, A. B. (2004) : Multidimensional variation for quasi-Monte Carlo.
- [11] SCHÖLKOPF, B. and SMOLA, A. J. (2002) : Learning with Kernels : Support Vector Machines, Regularisation, Optimization, and Beyond. Ed. The MIT Press.