

Fusion de détecteurs locaux par SVM. Application à la détection de ratés de combustion d'un moteur thermique

M. EZZEDDINNE¹, R. LENGELLE², A. BOATAS¹, N. LEFEBVRE¹

¹ PSA Peugeot-Citroën, 18 rue des Fauvelles, 92256 La Garenne Colombes

² ICD-LM2S, Université de technologie de Troyes, BP2060, 10010 Troyes cedex

Monjed.Ezzeddinne@mpsa.com, Regis.Lengelle@utt.fr, Armand.Boatas@mpsa.com, Nicolas.lefebvre@mpsa.com

Résumé – Ce papier présente une méthode de combinaison de détecteurs locaux par Machines à Vecteurs supports initialement proposée par Kwok. Une modification de la statistique de fusion conduit à un problème d'apprentissage de SVM classique. Ceci permet l'utilisation de divers algorithmes d'apprentissage utilisables pour des problèmes de grande taille. Cette méthode est appliquée dans le cas de la détection de ratés de combustion pour un moteur à essence.

Abstract – This paper presents a combination method of local detectors using Support Vector Machines initially proposed by Kwok. A modification of the mixture statistics leads to a classical formulation of SVM training. This allows the use of well known learning algorithms for large scale problems. This method is applied to ignition default detection for gasoline engines.

1. Introduction

La fusion de modèles a été proposée par Jacob et Jordan dans [4] et [6] pour des problèmes de régression et d'interpolation, dans des espaces de grande dimension. Les modèles locaux sont alors élaborés dans des sous espaces adaptés.

Les machines à vecteurs supports (SVM) sont particulièrement connues pour leurs capacités de généralisation, résultant de la minimisation du risque structurel [10][11]. Elles ont été introduites par Kwok [7] pour la fusion de modèles. Ce dernier a montré que le problème d'optimisation résultant est un problème quadratique sous contraintes relativement similaire au problème d'optimisation des SVM.

Dans cette version de l'algorithme de fusion, l'expression de la statistique proposée conduit à une solution approchée (obtenue par moindres carrés) pour l'estimation d'un certain terme de biais vectoriel. Cette solution approchée pose aussi le problème de la définition d'un critère d'arrêt d'un algorithme (itératif) d'apprentissage (voir partie 3).

De plus, une différence importante existe au niveau des contraintes. Celle-ci rend impossible l'utilisation efficace des algorithmes de sélection du working set et donc des algorithmes d'optimisation par décomposition.

Nous proposons ici une modification de la statistique de fusion qui permet de transformer le problème initial, mal posé, en un problème tel que :

- l'apprentissage est tout à fait comparable à celui d'une SVM standard (seul le noyau est modifié)
- les algorithmes efficaces de sélection de working set [9], [5] sont directement utilisables, ce qui

permet l'utilisation directe des méthodes par décomposition.

Ce papier est organisé comme suit. Dans la partie suivante, nous rappelons le problème d'optimisation quadratique sous contraintes associé à l'apprentissage des SVM en classification. La partie 3 est consacrée à l'algorithme de fusion par SVM initialement proposé par Kwok. La modification proposée est présentée en partie 4. La méthode est ensuite appliquée à la détection de ratés de combustion d'un moteur thermique en partie 5, suivie par une conclusion générale.

2. SVM pour la Classification

En classification non linéaire par SVM, la formulation duale du problème d'optimisation est donnée par :

$$\text{Min } \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^n \alpha_i \quad (\text{DP1})$$

avec :
$$\begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i \in \{1, \dots, n\} \end{cases}$$

α_i sont les multiplicateurs de Lagrange, chacun étant associé à une entrée $\mathbf{X}_i, i=1 \dots N$, où N est le nombre d'observations disponibles pour l'apprentissage.

La résolution d'un tel problème d'optimisation quadratique peut conduire à des problèmes de saturation de l'espace mémoire pour des grands jeux de données. Pour résoudre ces problèmes, des méthodes dites « par décomposition », reposant sur l'optimisation successive

d'un nombre réduit de multiplicateurs de Lagrange α_i , ont été proposées [1][5][8].

3. Fusion des détecteurs par SVM

Nous abordons ici l'utilisation des SVM pour la fusion de détecteurs locaux. Cette méthode, initialement proposée par Kwok [7], repose sur la statistique suivante :

$$f(\mathbf{X}) = \sum_{k=1}^P \pi_k(\mathbf{X}) \pi'_k(\mathbf{X})$$

où P est le nombre de détecteurs locaux, $\pi_k(\mathbf{X})$ est la sortie du $k^{\text{ème}}$ détecteur local (déterminé a priori), $\pi'_k(\mathbf{X})$ est une fonction d'activation à déterminer qui a pour objectif de combiner les détecteurs locaux de manière « optimale ». En restant dans le formalisme des SVM, nous écrivons :

$$\pi'_k(\mathbf{X}) = \langle \omega_k, \boldsymbol{\varphi}(\mathbf{X}) \rangle + \beta_k$$

La statistique globale de détection est donc la suivante :

$$f(\mathbf{X}) = \sum_{k=1}^P \pi_k(\mathbf{X}) (\langle \omega_k, \boldsymbol{\varphi}(\mathbf{X}) \rangle + \beta_k) \quad E_1$$

où ω_k , β_k sont des paramètres à déterminer pendant le processus d'apprentissage. Conformément au formalisme des SVM, le critère à optimiser est la somme d'un terme de maximisation de marge $\frac{1}{2} \sum_{k=1}^P \|\omega_k\|^2$ et d'un terme d'erreur $C \sum_{i=1}^N \xi_i$. Les contraintes sont tout à fait semblables à celles des SVM standards :

$$\begin{aligned} - y_i f(\mathbf{X}_i) &\geq 1 - \xi_i, i \in \{1, \dots, N\} \\ - \xi_i &\geq 0, i \in \{1, \dots, N\} \end{aligned}$$

La formulation duale du problème d'optimisation s'écrit alors :

$$\text{Min} \frac{1}{2} \sum_{k=1}^P \sum_{l=1}^P \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \pi_k(\mathbf{X}_i) \pi_k(\mathbf{X}_j) K(\mathbf{X}_i, \mathbf{X}_j) - \sum_{i=1}^N \alpha_i$$

$$\text{avec} \begin{cases} \sum_{i=1}^N \alpha_i y_i \pi_k(\mathbf{X}_i) = 0, & k \in \{1, \dots, P\} \\ 0 \leq \alpha_i \leq C, & i \in \{1, \dots, N\} \end{cases} \quad (\text{DP2})$$

$$\text{Posons } H(\mathbf{X}_i, \mathbf{X}_j) = \sum_{k=1}^P \pi_k(\mathbf{X}_i) \pi_k(\mathbf{X}_j) K(\mathbf{X}_i, \mathbf{X}_j).$$

D'après [3], H est un noyau vérifiant les conditions de Mercer. Le problème dual (DP2) s'écrit :

$$\text{Min} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j H(\mathbf{X}_i, \mathbf{X}_j) - \sum_{i=1}^N \alpha_i$$

$$\text{avec} \begin{cases} \sum_{i=1}^N \alpha_i y_i \pi_k(\mathbf{X}_i) = 0, & k \in \{1, \dots, P\} \\ 0 \leq \alpha_i \leq C, & i \in \{1, \dots, N\} \end{cases} \quad (\text{DP3})$$

(DP3) est un problème d'optimisation quadratique similaire à (DP1).

La résolution de ce problème permet l'identification des α_i et donc des paramètres ω_k , $k \in \{1, \dots, P\}$. Il convient maintenant de déterminer les β_k qui sont les termes de biais apparaissant dans la statistique globale de décision (E1) en exploitant les conditions de Karush Kuhn et Tucker (KKT). Celles-ci définissent les équivalences suivantes :

$$0 < \alpha_i < C \Leftrightarrow \text{Les individus sont sur la marge} \quad (\text{KKT1})$$

$$\alpha_i = 0 \Leftrightarrow \text{Les individus sont bien classés} \quad (\text{KKT2})$$

$$\alpha_i = C \Leftrightarrow \text{Les individus sont mal classés} \quad (\text{KKT3})$$

En réécrivant l'équation (KKT1), associée aux vecteurs supports qui se trouvent sur la marge, on obtient un système linéaire d'équations reliant le biais vectoriel aux erreurs du modèle (correspondant à un biais nul) par rapport aux données d'apprentissage, de la forme $\Delta \mathbf{Y} = \mathbf{H} \boldsymbol{\beta}$. Ce système n'admet, en général, pas de solution (autant d'équations que de vecteurs supports, P inconnues). Les β_k sont déterminés par moindres carrés. Une telle solution n'est pas exacte et introduit des perturbations sur l'interprétation des conditions KKT dans le problème primal. Leur application comme critère d'arrêt est donc impossible pour les méthodes par décomposition.

A ce problème s'ajoute une différence dans l'expression des contraintes du problème (DP3) par rapport au problème (DP1), qui interdit l'utilisation de certains algorithmes très efficaces de sélection du working set tel celui de Joachims [5].

4. Fusion avec biais scalaire

Afin de résoudre les problèmes identifiés ci-dessus, la statistique de fusion des détecteurs est modifiée comme suit :

$$f(\mathbf{X}) = \sum_{k=1}^P \pi_k(\mathbf{X}) \pi'_k(\mathbf{X}) + b \quad E_2$$

La restriction de la classe des fonctions de fusion sera, comme nous le constaterons expérimentalement, compensée par l'existence d'une solution exacte au problème d'optimisation. La fonction d'activation $\pi'_k(\mathbf{X})$ est modifiée en éliminant son biais explicite : $\pi'_k(\mathbf{X}) = \langle \omega_k, \boldsymbol{\varphi}(\mathbf{X}) \rangle$. La statistique globale de détection devient :

$$f(\mathbf{X}) = \left(\sum_{k=1}^P \pi_k(\mathbf{X}) \langle \omega_k, \boldsymbol{\varphi}(\mathbf{X}) \rangle \right) + b$$

La formulation duale du problème d'optimisation est :

$$\begin{aligned} & \text{Min } \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j H(\mathbf{X}_i, \mathbf{X}_j) - \sum_{i=1}^N \alpha_i \\ & \text{avec } \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i \in \{1, \dots, N\} \end{cases} \end{aligned} \quad (\text{DP4})$$

$$\text{où } H(\mathbf{X}_i, \mathbf{X}_j) = \sum_{k=1}^P \pi_k(\mathbf{X}_i) \pi_k(\mathbf{X}_j) K(\mathbf{X}_i, \mathbf{X}_j)$$

Le problème d'optimisation DP4 est strictement identique à DP1, seul le noyau est modifié. La solution résultante est une solution exacte, et les conditions KKT peuvent donc être exploitées afin de définir un critère d'arrêt pour les méthodes par décomposition.

5. Application à la détection de ratés de combustion

La détection de ratés de combustion est une fonction imposée par la législation « EOBD » (European On-Board Diagnostics) s'appliquant aux moteurs thermiques à essence qui exige qu'une surveillance en continu soit réalisée sur tous les véhicules automobiles de moins de 3,5T homologués à partir du 1^{er} janvier 2000. Tous les véhicules PSA Peugeot Citroën actuellement commercialisés en Europe sont donc équipés de cette fonctionnalité. Les méthodes de traitement du signal actuellement appliquées en temps réel sur véhicule sont reconnues globalement efficaces sur la zone spécifiée par la législation (du régime de ralenti à 4500 tr/mn). L'intérêt de cette surveillance est double. En premier lieu, elle permet de prendre les mesures permettant de limiter l'émission d'hydrocarbures imbrulés. En second lieu, elle permet d'éviter une élévation importante de la température du pot catalytique pouvant aller jusqu'à sa destruction.

Le signal principal utilisé pour la détection des ratés de combustion est le signal de régime « instantané » mesuré par le temps de passage de dents usinées sur le volant moteur devant une cible inductive. Chaque demi-tour entre l'angle correspondant au Point Mort Haut (PMH) d'un piston et son Point Mort Bas (PMB) correspond à la phase de détente d'un cylindre).

Un raté de combustion est caractérisé par la non-production de couple suite à une absence de combustion. On s'aperçoit alors que, sur le demi-tour correspondant à la phase de détente (signal de 30 échantillons correspondant à 30 durées de dents ; ce signal associé au couple (régime, charge) définit une « observation »), le signal de régime instantané change de forme et, dans les cas les plus flagrants, conduit à une diminution de la vitesse moyenne (augmentation de la durée de passage des dents).

La difficulté principale réside alors dans le fait que la forme du signal dépend fortement du point de fonctionnement du moteur : le régime et le remplissage (masse d'air admise dans le cylindre). Il faut donc proposer un modèle de décision construit par apprentissage qui, à partir des informations de régime instantané, du régime moyen et du remplissage (autrement dit la charge moteur), indique si le cylindre en phase de détente a brûlé son mélange carburé ou non. Par ailleurs, une altération de la vitesse instantanée de rotation est observée sur plusieurs rotations consécutives du moteur, en raison du couplage du moteur et de la chaussée via l'ensemble de la chaîne cinématique (cf. FIG. 1).

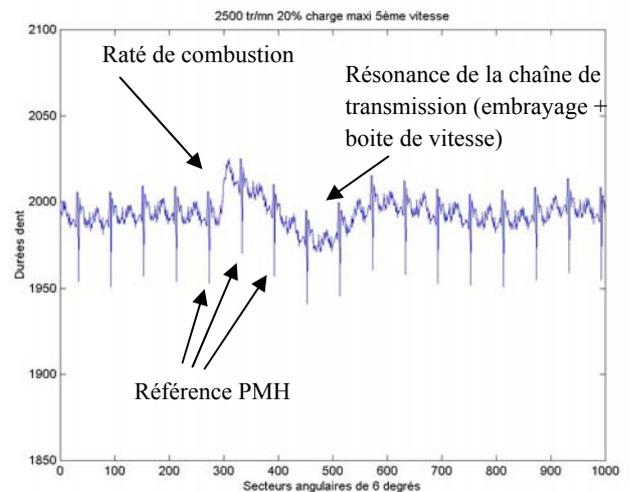


FIG. 1 : Exemple de signal de durée dents (faible régime, faible charge)

Ainsi qu'il a été dit plus haut, l'altération du signal de vitesse instantanée est fonction des paramètres de fonctionnement du moteur (régime, charge). Le détecteur « optimal » sera donc fonction de ces paramètres. On observe sur la figure 2 que le raté de combustion se traduit par une très faible modification du signal de durée dent.

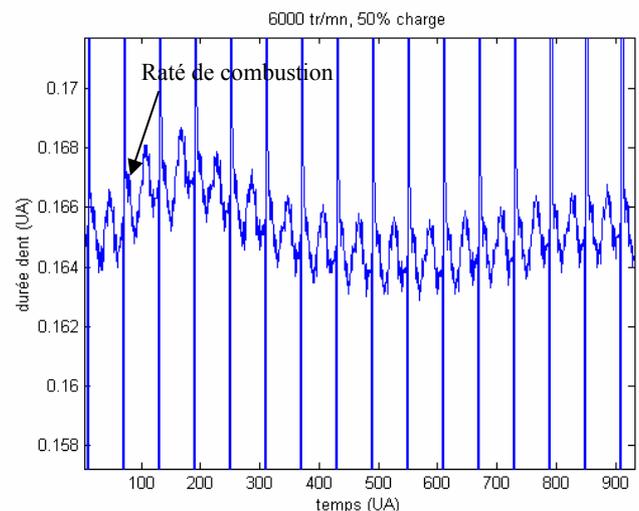


FIG. 2 : Exemple de signal de durée dents (régime élevé, charge moyenne)

La procédure expérimentale retenue est la suivante. Pour chaque valeur du régime moteur (de 1000 à 6000 tr/mn par pas de 250 tr/mn), pour chaque valeur de la charge (20, 35, 50, 65, 80 % de la charge maximale), pour les différents rapports de boîte et pour différentes fréquences de génération des ratés (ces deux derniers paramètres jouent sur l'excitation de la chaîne cinématique), les signaux de mesure de durée de passage des dents ont été enregistrés ainsi qu'un signal binaire indiquant la présence/l'absence de la génération d'un raté de combustion.

Pour chaque valeur du couple (régime, charge), un détecteur polynomial a été optimisé (Support Vector Machine). Une procédure de validation croisée a permis de sélectionner l'ordre 1 pour chaque détecteur local (donc linéaire). La fusion des détecteurs locaux a été réalisée conformément à la méthode présentée dans ce papier.

Les performances globales (probabilité de non détection, probabilité de fausse alarme) ont été évaluées par une approche de type « Leaving One Out ». Les données de chaque couple (régime charge) ont été supprimées itérativement de l'ensemble d'apprentissage lors de l'étape de fusion des détecteurs locaux. La fusion a été réalisée sur les autres données disponibles et les données supprimées ont été utilisées comme données de test.

La figure 3, courbe du haut, représente la valeur de la statistique de détection du détecteur local en fonction du n° de l'observation (les vrais ratés ont été mis en première partie de courbe (valeur $\approx +1$). Les non ratés sont en seconde partie de courbe (valeur ≈ -1)). La courbe inférieure représente, dans les mêmes conditions, le résultat de la fusion des détecteurs locaux. Pour ce point de fonctionnement du moteur, les probabilités estimées de non détection et de fausse alarme sont respectivement nulle et de l'ordre de 10^{-3} .

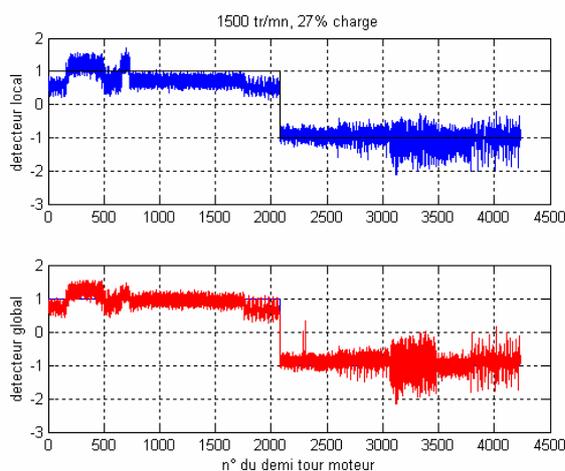


FIG. 3 : Test du modèle de fusion

6. Conclusion

Dans ce papier, nous avons proposé une méthode de combinaison de détecteurs locaux dans un formalisme

SVM. La formulation proposée conduit à un problème d'optimisation quadratique sous contraintes tout à fait semblable à celui de l'apprentissage de SVM en classification. Notre méthode a été ensuite appliquée à la détection de ratés de combustion. Le détecteur repose sur deux niveaux de traitement. Le premier correspond à la détection conditionnée par chaque point de fonctionnement (régime, charge) disponible dans les données fournies. Le second consiste à « fusionner » ces détecteurs individuels afin d'obtenir une statistique de détection unique. La fusion conduit à un niveau de performances tout à fait satisfaisant, associé à une grande robustesse vis-à-vis des paramètres d'analyse, qui sont, de plus, peu nombreux. Quelques résultats pour la détection des ratés de combustion ont été présentés en cinquième partie. Cette méthode a été étendue au cas de la fusion de régresseurs locaux pour l'estimation des émissions de polluants d'un moteur thermique [2].

Références

- [1] R. Collobert et S. Bengio, *SVM Torch. Support Vector Machine for Large-Scale Regression and Classification Problems*. Journal of Machine Learning Research, 1: pp. 143-160, 2001.
- [2] M. Ezzeddinne, R. Lengellé, *Mixture of local models for pollutants emission modeling*, EAEC 2007, Budapest, 30 mai-1er juin 2007. Sur CDROM.
- [3] M. Genton. *Classes of Kernels for Machine Learning: A statistics Perspective*. Journal of Machine Learning Research 2: pp. 299-312, 2001.
- [4] R. Jacobs, M. Jordan, S. Nowlan et G. Hinton. *Adaptive Mixtures of Local Experts*. Neural Computation, 3(1): pp. 79-87, 1991.
- [5] T. Joachims, Making. *Large-Scale SVM Learning Practical*. Advances in Kernel Methods – Support Vector Learning, ch. 11, pp. 169-184, MIT Press, 1999.
- [6] M. Jordan et R. Jacobs. *Hierarchical Mixtures of Experts and the EM Algorithm*. Neural Computation, 6(2): pp. 181-214, 1994.
- [7] J. Kwok. *Support Vector Mixture for Classification and Regression Problems*. Proceedings of the International Conference on Pattern Recognition, pp. 255-258, 1998.
- [8] C. J. Lin. *On the Convergence of the Decomposition Method for Support Vector Machines*. IEEE Transactions on Neural Networks, 12 (2001), pp. 1288-1298.
- [9] E. Osuna, R. Freund et F. Girosi. *An Improved Training Algorithm for Support Vector Machines*. Proceedings of IEEE Workshop on Neural Networks for Signal Processing pp. 276-285, 1997.
- [10] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [11] V. Vapnik. *Statistical Learning Theory*. A Wiley-Interscience Publication, New York, 1998.